Europäisches Patentamt

European Patent Office

Office européen des brevets



(11) EP 0 880 590 B1

(12)

EUROPEAN PATENT SPECIFICATION

- (45) Date of publication and mention of the grant of the patent:
 20.10.2004 Bulletin 2004/43
- (21) Application number: 97906528.1
- (22) Date of filing: 11.02.1997

- (51) Int Cl.7: **C12N 15/55**, C12N 9/16, C12N 1/21, C12P 13/02, C12Q 1/68, C07K 16/40, G01N 33/53
- (86) International application number: PCT/US1997/002039
- (87) International publication number: WO 1997/030160 (21.08.1997 Gazette 1997/36)

(54) ESTERASES

ESTERASEN

ESTERASES

- (84) Designated Contracting States:

 AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC

 NL PT SE
- (30) Priority: 16.02.1996 US 602359
- (43) Date of publication of application: 02.12.1998 Bulletin 1998/49
- (60) Divisional application: 04017203.3
- (73) Proprietor: DIVERSA CORPORATION San Diego, California 92121 (US)
- (72) Inventors:
 - ROBERTSON, Dan, E. Haddonfield, NJ 08033 (US)
 - MURPHY, Dennis Paoli, PA 19301 (US)
 - REID, John Bryn Mawr, PA 19010 (US)
 - MAFFIA, Anthony, M.
 Wilmington, DE 19810 (US)
 - LINK, Steven
 Wilmington, DE 19803 (US)
 - SWANSON, Ronald, V. Media, PA 19063 (US)
 - WARREN, Patrick, V.
 Philadelphia, PA 19163 (US)
 - KOSMOTKA, Anna Brookhaven, PA 19015 (US)
 - CALLEN, Walter San Diego, CA 92117 (US)

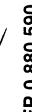
- (74) Representative: VOSSIUS & PARTNER Siebertstrasse 4 81675 München (DE)
- (56) References cited:

EP-A- 0 318 775 WO-A-94/02617 WO-A-93/25691 WO-A-95/23853

WO-A-95/34644

US-A- 5 306 636

- WRAY L. ET AL.: "Streptomyces coelicolor transcription factor, AC L03213" EBI DATABASE,9 October 1992 (1992-10-09), XP002145895
- KUGIMIYA WATARU ET AL: "Molecular cloning and structure of the gene for esterase from a thermophilic bacterium, Bacillus stearothermophilus IFO 12550." BIOSCIENCE BIOTECHNOLOGY AND BIOCHEMISTRY 1992, vol. 56, no. 12, 1992, pages 2074-2075, XP002138593 ISSN: 0916-8451
- HUDDLESTON S ET AL: "The identification and partial characterisation of a novel inducible extracellular thermostable esterase from the archaeon Sulfolobus shibatae." BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS 1995, vol. 216, no. 2, 1995, pages 495-500, XP002138594 ISSN: 0006-291X



Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Printed by Jouve, 75001 PARIS (FR)

- DERWENT BIOTECHNOLOGY ABSTRACTS, Vol. 14, No. 15, 17 July 1995, page 125, Abstract No. 95-09223, LADRAT C. et al., "Thermostable Enzymes Screened on Thermophilic Microorganisms from Deep-Sea Hydrothermal Vents - -beta- Glucosidase, Alcohol Dehydrogenase, Protease and Esterase Activity"; & C.R. SEANCES ACAD. SCI. SER 3., 1995, 318(4), 423-429.
- DERWENT BIOTECHNOLOGY ABSTRACTS, Vol. 13, No. 15, 25 July 1994, pages 95-96, Abstract No. 94-08869, CORNEC L. et al., "Thermostable Esterase Activity from Deep-Sea Sulfur Dependent Thermophilic Archaebacteria-Thermostable Enzyme Purification from Thermophilic Archaebacterium and Characterization"; & THERMOPHILES SCI. TECHNOL., 1992, 18.

 APPL. MICROBIOL. BIOTECHNOL., 1994, Vol. 42, CANGANELLA F., "Characterization of Amylolytic and Pullulytic Enzymes from Thermophilic Archea and from a New Fervidobacterium Species", pages 239-245.

Description

40

[0001] This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polynucleotides and polypeptides of the present invention have been putatively identified as esterases. Esterases are enzymes that catalyze the hydrolysis of ester groups to organic acids and alcohols.

[0002] Many esterases are known and have been discovered in a broad variety of organisms. including bacteria. yeast and higher animals and plants. A principal example of esterases are the lipases, which are used in the hydrolysis of lipids, acidolysis (replacement of an esterified fatty acid with a free fatty acid) reactions, transesterification (exchange of fatty acids between triglycerides) reactions, and in ester synthesis. The major industrial applications for lipases include: the detergent industry, where they are employed to decompose fatty materials in laundry stains into easily removable hydrophilic substances: the food and beverage industry where they are used in the manufacture of cheese, the ripening and flavoring of cheese, as antistaling agents for bakery products, and in the production of margarine and other spreads with natural butter flavors; in waste systems; and in the pharmaceutical industry where they are used as digestive aids.

[0003] The polynucleotides and polypeptides of the present invention have been identified as esterases as a result of their enzymatic activity.

[0004] In accordance with one aspect of the present invention, there are provided novel enzymes. as well as active fragments, analogs and derivatives thereof.

[0005] In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding the enzymes of the present invention including mRNAs, cDNAs, genomic DNAs as well as active analogs and fragments of such enzymes.

[0006] In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence of the present invention, under conditions promoting expression of said enzymes and subsequent recovery of said enzymes.

[0007] In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes for hydrolyzing ester groups to yield an organic acid and an alcohol. The esterases of the invention are stable at high temperatures and in organic solvents and, thus, are superior for use in production of optically pure chiral compounds used in pharmaceutical, agricultural and other chemical industries.

[0008] In accordance with yet a further aspect of the present invention, there are also provided nucleic acid probes comprising nucleic acid molecules of sufficient length to hybridize to a nucleic acid sequence of the present invention.

[0009] In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes, for *in vitro* purposes related to scientific research, for example, to generate probes for identifying similar sequences which might encode similar enzymes from other organisms by using certain regions, *i.e.*, conserved sequence regions, of the nucleotide sequence.

[0010] These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

[0011] The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

[0012] Figure 1 is an illustration of the full-length DNA (SEQ ID NO:23) and corresponding deduced amino acid sequence (SEQ ID NO:33) of *Staphylothermus marinus* F1-12LC of the present invention. Sequencing was performed using a 378 automated DNA sequencer (Applied Biosystems, Inc.) for all sequences of the present invention.

[0013] Figure 2 is an illustration of the full-length DNA (SEQ ID NO:24) and corresponding deduced amino acid sequence (SEQ ID NO:34) of *Pyrodictium* TAG 11-17LC.

[0014] Figure 3 is an illustration of the full-length DNA (SEQ ID NO:25) and corresponding deduced amino acid sequence (SEQ ID NO:35) of *Archaeoglobus venificus* SNP6-24LC.

[0015] Figure 4 is an illustration of the full-length DNA (SEQ ID NO:26) and corresponding deduced amino acid sequence (SEQ ID NO:36) of *Aquifex pyrophilus*-28LC.

[0016] Figure 5 is an illustration of the full-length DNA (SEQ ID NO:27) and corresponding deduced amino acid sequence (SEQ ID NO:37) of M11TL-29L.

[0017] Figure 6 is an illustration of the full-length DNA (SEQ ID NO:28) and corresponding deduced amino acid sequence (SEQ ID NO:38) of *Thermococcus* CL-2-30LC.

[0018] Figure 7 is an illustration of the full-length DNA (SEQ ID NO:29) and corresponding deduced amino acid sequence (SEQ ID NO:39) of Aquifex VF5-34LC.

[0019] Figure 8 is an illustration of the full-length DNA (SEQ ID NO:30) and corresponding deduced amino acid

sequence (SEQ ID NO:40) of Teredinibacter-42L.

[0020] Figure 9 is an illustration of the full-length DNA (SEQ ID NO:31) and corresponding deduced amino acid sequence (SEQ ID NO:41) of *Archaeoglobus fulgidus VC*16-16MC.

[0021] Figure 10 is an illustration of the full-length DNA (SEQ ID NO:32) and corresponding deduced amino acid sequence (SEQ ID NO:42) of *Sulfolobus solfataricus* P1-8LC.

[0022] Figure 11 is an illustration of the full-length DNA (SEQ ID NO:43) and corresponding deduced amino acid sequence (SEQ ID NO:53) of LA11.1 Esterase es2 of the present invention.

[0023] Figure 12 is an illustration of the full-length DNA (SEQ ID NO:44) and corresponding deduced amino acid sequence (SEQ ID NO:54) of Whale Mat Sample 11.801 Esterase es9.

[0024] Figure 13 is an illustration of the full-length DNA (SEQ ID NO:45) and corresponding deduced amino acid sequence (SEQ ID NO:55) of *Metallosphaera Prunae Ron* 12/2 Esterase 23mc1.

[0025] Figure 14 is an illustration of the full-length DNA (SEQ ID NO:46) and corresponding deduced amino acid sequence (SEQ ID NO:56) of *Thermotoga. neapolitana* 5068 Esterase 56mc4.

[0026] Figure 15 is an illustration of the full-length DNA (SEQ ID NO:47) and corresponding deduced amino acid sequence (SEQ ID NO:57) of *Melittangium lichenicola* Esterase 77mc1.

[0027] Figure 16 is an illustration of the full-length DNA (SEQ ID NO:48) and corresponding deduced amino acid sequence (SEQ ID NO:58) of Whale Mat Sample 11.801 Esterase es2.

[0028] Figure 17 is an illustration of the full-length DNA (SEQ ID NO:49) and corresponding deduced amino acid sequence (SEQ ID NO:59) of Whale Mat Sample AD3059 Esterase es4.

[0029] Figure 18 is an illustration of the full-length DNA (SEQ ID NO:50) and corresponding deduced amino acid sequence (SEQ ID NO:60) of *Microscilla furvescens* Esterase 53sc2.

[0030] figure 19 is an illustration of the full-length DNA (SEQ ID NO:51) and corresponding deduced amino acid sequence (SEQ ID NO:61) of *Thermotoga maritima MSB8* Esterase 6sc1.

[0031] Figure 20 is an illustration of the full-length DNA (SEQ ID NO:52) and corresponding deduced amino acid sequence (SEQ ID NO:62) of *Polyangium brachysporum* Esterase 78mc1.

[0032] The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

[0033] A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences. The coding sequences need not be contiguous to one another so long as the expressed sequences ultimately process to produce the desired protein.

[0034] "Recombinant" enzymes refer to enzymes produced by recombinant DNA techniques; *i.e.*, produced from cells transformed by an exogenous DNA construct encoding the desired enzyme. "Synthetic" enzymes are those prepared by chemical synthesis.

[0035] A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular enzyme, is a DNA sequence which is transcribed and translated into an enzyme when placed under the control of appropriate regulatory sequences. [0036] In accordance with an aspect of the present invention, there are provided isolated nucleic acids (polynucleotides) which encode for the mature enzymes having the deduced amino acid sequences of Figures 1-20 (SEQ ID NOS:23-32 and 43 to 52.)

[0037] The deposit(s) have been made under the terms of the Budapest Treaty on the International Recognition of the deposit of micro-organisms for purposes of patent procedure. The strains will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit would be required under 35 U.S.C. §112. The sequences of the polynucleotides contained in the deposited materials, as well as the amino acid sequences of the polypeptides encoded thereby, are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

[0038] The polynucleotides of this invention were originally recovered from genomic gene libraries derived from the following organisms:

6 [0039] Staphylothermus marinus F1 is a thermophilic sulfur archaea which was isolated in Vulcano, Italy. It grows optimally at 85°C (T_{max} = 98°C) at pH 6.5.

[0040] Pyrodictium TAG11 is a thermophilic sulfur archaea which was isolated in the Middle Atlantic Ridge. It grows optimally at 103° C ($T_{max} = 110^{\circ}$ C) at pH 6.5.

[0041] Archaeoglobus venificus SNP6 was isolated in the Middle Atlantic Ridge and grows optimally at 75°C (T_{max} = 92°C) at pH 6.9.

[0042] Aquifex pyrophilus K01 5a was isolated at Kolbeinsey Ridge, North of Iceland. This marine organism is a gram-negative, rod-shaped, strictly chemolithoautrophic, knall gas bacterium. It grows optimally at 85°C (T_{max} = 95°C) at pH 6.8.

[0043] M11TL is a new species of Desulfurococcus which was isolated from Diamond Pool (formerly Jim's Black Pool) in Yellowstone. The organism grows heterotrophically by fermentation of different organic materials (sulfur is not necessary) in grape-like aggregates optimally at 85 - 88°C in a low salt medium at pH 7.0.

[0044] Thermococcus CL-2 was isolated in the North Cleft Segment of the Juan de Fuca Ridge from a severed alvinellid worm residing on a "black smoker" sulfide structure. This marine archaea forms pleomorphic cocci, and grows optimally at 88°C.

[0045] Aquifex VF5 was isolated at a beach in Vulcano, Italy. This marine organism is a gram-negative, rod-shaped, strictly chemolithoautotrophic, knall gas bacterium. It grows optimally at 85°C (T_{max} = 95°C) at pH 6.8.

[0046] Teredinibacter (pure) is an endosymbiont of the shipworm Bankia gouldi. The organism has straight to slightly bent 5-10 µm rods, and forms spiral cells as stationary phase is met. The organism was described in Science (1983) 22:1401-1403. It grows optimally at 30°C at pH 8.0.

[0047] Archaeoglobus fulgidus VC16 was isolated in Vulcano, Italy. The organism grows optimally at 85°C (T_{max} = 92°C) at pH 7.0.

[0048] Sulfolobus solfataricus P1 grows optimally at 85°C ($T_{max} = 87$ °C) at pH 2.0.

[0049] Accordingly, the polynucleotides and enzymes encoded thereby are identified by the organism from which they were isolated, and are sometimes hereinafter referred to as F1/12LC (Figure 1 and SEQ ID NOS:23 and 33), TAG11/17LC (Figure 2 and SEQ ID NOS:24 and 34), SNP6/24LC (Figure 3 and SEQ ID NOS:25 and 35), AqP/28LC (Figure 4 and SEQ ID NOS:26 and 36), M11TL/29L (Figure 5 and SEQ ID NOS:27 and 37), CL-2/30LC (Figure 6 and SEQ ID NOS:28 and 38), VF5/34LC (Figure 7 and SEQ ID NOS:29 and 39), Trb/42L (Figure 8 and SEQ ID NOS:30 and 40). VC16/16MC (Figure 9 and SEQ ID NOS:31 and 41) and P1/8LC (Figure 10 and SEQ ID NOS: 32 and 42).

[0050] The polynucleotides and polypeptides of the present invention show identity at the nucleotide and protein level to known genes and proteins encoded thereby as shown in Table 1.

Table 1

			Table I		
25	Enzyme	Gene w/closest Homology (Organism)	Protein Similarity (%)	Protein Identity (%)	DNA Identity (%)
	F1/12LC	No significant homology	-	-	
	TAG11/17LC	No significant homology	<u>.</u>	-	-
30	SNP6/24LC	PIR S34609 - carboxylesterase Pseudomones sp. (strain KWI- 56) open reading frame of unknown function in E.coli.	46	27	42
	AqP/29LC		53	31	38
	M11TL/29LC	No significant homology	-	-	-
	CL02/30LC	No significant homology	•	-	-
40	VF5/34LC	Identified by homology to 28LC; also homologous to ORF of unknown function 5' of tgs in <i>E. coli</i>	84	71	71
45	Trb/42L	No significant homology	-	-	-
	P1-8LC		-		
	VC16-16MC				

[0051] All the clones identified in Table 1 encode polypeptides which have esterase activity.

[0052] This invention, in addition to the isolated nucleic acid molecules encoding the enzymes of the present invention, also provides substantially similar sequences. Isolated nucleic acid sequences are substantially similar if: (i) they are capable of hybridizing under conditions hereinafter described, to the polynucleotides of SEQ ID NOS:23-32 and 43-52; (ii) or they encode DNA sequences which are degenerate to the polynucleotides of SEQ ID NOS:23-32 and 43 to 52 Degenerate DNA sequences encode the amino acid sequences of SEQ ID NOS:33-42 and 53 to 62, but have variations in the nucleotide coding sequences. As used herein, substantially similar refers to the sequences having similar identity to the sequences of the instant invention. The nucleotide sequences that are substantially the same can be identified by hybridization or by sequence comparison. Enzyme sequences that are substantially the same can

be identified by one or more of the following: proteolytic digestion, gel electrophoresis and/or microsequencing.

[0053] One means for isolating the nucleic acid molecules encoding the enzymes of the present invention is to probe a gene library with a natural or artificially designed probe using art recognized procedures (see, for example: Current Protocols in Molecular Biology, Ausubel F.M. *et al.* (EDS.) Green Publishing Company Assoc. and John Wiley Interscience, New York, 1989, 1992). It is appreciated by one skilled in the art that the polynucleotides of SEQ ID NOS: 23-32 and 43 to 52 or fragments thereof (comprising at least 12 contiguous nucleotides), are particularly useful probes. Other particularly useful probes for this purpose are hybridizable fragments of the sequences of SEQ ID NOS:1-22 (*i. e.*, comprising at least 12 contiguous nucleotides).

[0054] With respect to nucleic acid sequences which hybridize to specific nucleic acid sequences disclosed herein, hybridization may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions. As an example of oligonucleotide hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9 M NaCl, 50 mM NaH₂PO₄, pH 7.0. 5.0 mM Na₂EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/mL polyriboadenylic acid. Approximately 2 X 10⁷ cpm (specific activity 4-9 X 10⁸ cpm/ug) of ³²P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in !X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na₂EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm -10°C for the oligo-nucleotide probe. The membrane is then exposed to autoradiographic film for detection of hybridization signals.

[0055] Stringent conditions means hybridization will occur only if there is at least 90% identity, preferably at least 95% identity and most preferably at least 97% identity between the sequences. See J. Sambrook et al., Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory (1989) which is hereby incorporated by reference in its entirety.

[0056] As used herein, a first DNA (RNA) sequence is at least 70% and preferably at least 80% identical to another DNA (RNA) sequence if there is at least 70% and preferably at lest a 80% or 90% identity, respectively, between the bases of the first sequence and the bases of the another sequence, when property aligned with each other, for example when aligned by BLASTN.

[0057] The present invention relates to polynucleotides which differ from the reference polynucleotide such that the changes are silent changes, for example the change do not after the amino acid sequence encoded by the polynucleotide. The present invention also relates to nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference polynucleotide. In a preferred aspect of the invention these polypeptides retain the same biological action as the polypeptide encoded by the reference polynucleotide.

[0058] The polynucleotides of this invention were recovered from genomic gene libraries from the organisms listed in Table 1. Gene libraries were generated in the Lambda ZAP II cloning vector (Stratagene Cloning Systems). Mass excisions were performed on these libraries to generate libraries in the pBluescript phagemid. Libraries were generated and excisions were performed according to the protocols/methods hereinafter described.

[0059] The polynucleotides of the present invention may be in the form of RNA or DNA which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequences which encodes the mature enzymes may be identical to the coding sequences shown in Figures 1-20 (SEQ ID NOS:23-32) and 43 to 52 or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, and 43-52, encodes the same mature enzymes as the DNA of Figures 1-20 (SEQ ID NOS:23-32 and 43-52).

[0060] The polynucleotide which encodes for the mature enzyme of Figures 1-20 (SEQ ID NOS:33-42) and 53-62 may include, but is not limited to: only the coding sequence for the mature enzyme; the coding sequence for the mature enzyme and additional coding sequence such as a leader sequence or a proprotein sequence; the coding sequence for the mature enzyme (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature enzyme.

[0061] Thus, the term "polynucleotide encoding an enzyme (protein)" encompasses a polynucleotide which includes only coding sequence for the enzyme as well as a polynucleotide which includes additional coding and/or non-coding sequence.

50

[0062] The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the enzymes having the deduced amino acid sequences of Figures 1-20 (SEQ ID NOS:33-42 and 53-62). The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

[0063] Thus, the present invention includes polynucleotides encoding the same mature enzymes as shown in Figures 1-20 (SEQ ID NOS:23-32 and 43-52) as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the enzymes of Figures 1-20 (SEQ ID NOS:23-32 and 43 to 52). Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

[0064] As hereinabove indicated, the polynucleotides may have a coding sequence which is a naturally occurring allelic variant of the coding sequences shown in Figures 1-20 (SEQ ID NOS:23-32 and 43 to 52). As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded enzyme.

[0065] Fragments of the full length gene of the present invention may be used as hybridization probes for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promoter regions, exons and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of genomic DNA to determine which members of the library the probe hybridizes to.

10

30

35

50

[0066] It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate complementary copies of DNA from other sources or to screen such sources for related sequences.

[0067] The present invention further relates to polynucleotides or fragments thereof encoding an enzyme having esterase activity and which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature enzyme encoded by the DNA of Figures 1-20 (SEQ ID NOS:23-32 and 43 to 52.)

[0068] Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases of a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For exarriple, such polynucleotides may be employed as probes for the polynucleotides of SEQ ID NOS:23-32, and 43 to 52 for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

[0069] Thus, the present invention is directed to polynucleotides or fragments thereof encoding an enzyme having esterase activity and having at least a 70% identity, preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes the enzymes of SEQ ID NOS:33-42 and 53 to 62 as well as fragments thereof, which fragments have at least 15 bases, preferably at least 30 bases and most preferably at least 50 bases. which fragmerits are at least 90% identical, preferably at least 95% identical and most preferably at least 97% identical under stringent conditions to any portion of a polynucleotide of the present invention.

[0070] The present invention further relates to enzymes which have the deduced amino acid sequences of Figures 1-20 (SEQ ID NOS:23-32 and 43 to 52) as well as fragments, analogs and derivatives of such enzyme.

[0071] The terms "fragment," "derivative" and "analog" when referring to the enzymes of Figures 1-20 (SEQ ID NOS: 33-42 and 53 to 62) mean enzymes which retain essentially the same biological function or activity as such enzymes. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature enzyme.

[0072] The enzymes of the present invention may be a recombinant enzyme, a natural enzyme or a synthetic enzyme, preferably a recombinant enzyme.

[0073] The fragment, derivative or analog of the enzymes of Figures 1-20 (SEQ ID NOS:33-42 and 53 to 62) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code. or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature enzyme is fused with another compound, such as a compound to increase the half-life of the enzyme (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature enzyme, such as a leader or secretory sequence or a sequence which is employed for purification of the mature enzyme or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

[0074] The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

[0075] The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials

in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

[0076] The enzymes of the present invention include the enzymes of SEQ ID NOS:33-42 and 53 to 62 (in particular the mature enzyme) as well as enzymes which have at least 70 % similarity (preferably at least 70% identity) to the enzymes of SEQ ID NOS:33-42 and 53 to 62 and more preferably at least 90% similarity (more preferably at least 90% identity) to the enzymes of SEQ ID NOS:33-42 and 53 to 62 and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the enzymes of SEQ ID NOS:33-42 and 53 to 62 and also include portions of such enzymes with such portion of the enzyme generally containing at least 30 amino acids and more preferably at least 50 amino acids.

[0077] As known in the art "similarity" between two enzymes is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one enzyme to the sequence of a second enzyme.

10

[0078] A variant, i.e. a "fragment", "analog" or "derivative" polypeptide, and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

[0079] Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

[0080] Most highly preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies.

[0081] Fragments or portions of the enzymes of the present invention may be employed for producing the corresponding full-length enzyme by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length enzymes. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

[0082] The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of enzymes of the invention by recombinant techniques.

[0083] Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

[0084] The polynucleotides of the present invention may be employed for producing enzymes by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing an enzyme. Such vectors include chromosomal nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus. fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

[0085] The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

[0086] The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli. lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

[0087] In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

[0088] The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

[0089] As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as E. coli.

Strepromyces, Bacillus subtilis; fungal cells, such as yeast; insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS or Bowes melanoma: adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

[0090] More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pBluescript II KS, ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia); Eukaryotic: pXT1, pSG5 (Stratagene) pSVK3, pBPV, pMSG, pSVL, SV40 (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

[0091] Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt. lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

[0092] In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

[0093] The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the enzymes of the invention can be synthetically produced by conventional peptide synthesizers.

25

[0094] Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokary-otic and eukaryotic hosts are described by Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, Second Edition*, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

[0095] Transcription of the DNA encoding the enzymes of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

[0096] Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of $E.\ coli$ and $S.\ cerevisiae$ TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK). α -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated enzyme. Optionally, the heterologous sequence can encode a fusion enzyme including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

[0097] Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli, Bacillus subtilis, Salmonella typhimurium* and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

[0098] As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. [0099] Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

[0100] Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

[0101] Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

[0102] Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell, 23*: 175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

[0103] The enzyme can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

[0104] The enzymes of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the enzymes of the present invention may be glycosylated or may be non-glycosylated. Enzymes of the invention may or may not also include an initial methionine amino acid residue.

[0105] Esterases are a group of key enzymes in the metabolism of fats and are found in all organisms from microbes to mammals. In the hydrolysis reaction, an ester group is hydrolysed to an organic acid and an alcohol.

[0106] Esterases enantiomerically differentiate dicarboxylic diesters and diacetates of diols. Using the approach disclosed in a commonly assigned, copending provisional application Serial No. 60/008,316, filed on December 7, 1995 and entitled "Combinatorial Enzyme Development," the disclosure of which is incorporated herein by reference in its entirety, one could convert the enantiospecificity of the esterase. Further, the thermostable esterases are believed to have superior stability at higher temperatures and in organic solvents. Thus, they are better suited for use in rigorous production procees which require robust catalysts.

[0107] There are a number of industrial and scientific applications for esterases, such as those of the present invention. including:

- 1) Esterases are useful in the dairy industry as ripening starters for cheeses, such as the Swiss-type cheeses;
- 2) Esterases are useful in the pulp and paper industry for lignin removal from cellulose pulps, for lignin solubilization by cleaving the ester linkages between aromatic acids and lignin and between lignin and hemicelluloses, and for disruption of cell wall structure when used in combination with xylanase and other xylan-degrading enzymes in biopulping and biobleaching of pulps;
- 3) Esterases are useful in the synthesis of carbohydrate derivatives, such as sugar derivatives;

35

40

45

- 4) Esterases are useful, when combined with xylanases and cellulases, in the conversion of lignocellulosic wastes to fermentable sugars for producing a variety of chemicals and fuels;
- 5) Esterases are useful as research reagents in studies on plant cell wall structure, particularly the nature of covalent bonds between lignin and carbohydrate polymers in the cell wall matrix;
- 6) Esterases are also useful as research reagents in studies on mechanisms related to disease resistance in plants and the process of organic matter decomposition; and
- 7) Esterases are useful in selection of plants bred for production of highly digestible animal feeds, particularly for ruminant animals.
- [0108] Antibodies generated against the enzymes corresponding to a sequence of the present invention can be obtained by direct injection of the enzymes into an animal or by administering the enzymes to an animal, preferably a nonhuman. The antibody so obtained will then bind the enzymes itself. In this manner, even a sequence encoding only a fragment of the enzymes can be used to generate antibodies binding the whole native enzymes. Such antibodies

can then be used to isolate the enzyme from cells expressing that enzyme.

[0109] For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, *Nature*, 256:495-497, 1975), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* 4:72, 1983), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.*, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96, 1985).

[0110] Techniques described for the production of single chain antibodies (U.S. Patent 4.946.778) can be adapted to produce single chain antibodies to immunogenic enzyme products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic enzyme products of this invention.

[0111] Antibodies generated against an enzyme of the present invention may be used in screening for similar enzymes from other organisms and samples. Such screening techniques are known in the art, for example, one such screening assay is described in Sambrook *et al.*, Molecular Cloning: A Laboratory Manual (2d Ed.), Cold Spring Harbor Laboratory, Section 12.21-12.28 (1989) which is hereby incorporated by reference in its entirety.

[0112] The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

[0113] In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

[0114] "Plasmids" are designated by a lower case "p" preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

[0115] "Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 μ g of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μ l of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μ g of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37' C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

[0116] Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel et al., Nucleic Acids Res., 8:4057 (1980).

[0117] "Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

[0118] "Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

[0119] Unless otherwise stated. transformation was performed as described in Sambrook *et al.*, Molecular Cloning: A Laboratory Manual (2d Fd.), Cold Spring Harbor Press (1989).

45 Example 1

Bacterial Expression and Purification of Esterases

[0120] DNA encoding the enzymes of the present invention, SEQ ID NOS:33 through 42, were initially amplified from a pBluescript vector containing the DNA by the PCR technique using the primers noted herein. The amplified sequences were then inserted into the respective PQE vector listed beneath the primer sequences. and the enzyme was expressed according to the protocols set forth herein. The 5' and 3' primer sequences for the respective genes are as follows:

Staphylothermus marinus F1-12LC

55

5' CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGTCTTTA AACAAGCACT CT

	3' CGGAAGATCT CTATCGTTTA GTGTATGATT T
5	vector: pQET
	Pyrodictium TAG 11-17LC
10	5' CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGAAACTC CTTGAGCCCA CA 3' CGGAAGATCT CGCCGGTACA CCATCAGCCA C Bg1
15	vector: pQET
	Archaeoglobus venificus SNP6-24LC
20	5' CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCCATAT GTTAGGAATG GT 3' CGGAGGTACC TTAGAACTGT GCTGAAGAAA TAAATTCGTC CATTGCTCTA TTA
25	vector: pQET
	Aquifex pyrophilus - 28LC
30	5' CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGAGATTG AGGAAATTTG AAG 3' CGGAGGTACC CTATTCAGAA AGTACCTCTA A
35	vector: pQET
	M11TL - 29LC
40	5' CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGTTTAAT ATCAATGTCT TT 3' CGGAAGATCT TTAAGGATTT TCCCTGGGTA G
	vector: pQET
45	Thermococcus CL-2 - 30LC
50	5' CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGGAGGTT TACAAGGCCA AA 3' CGGAGGTACC TTATTGAGCC GAAGAGTACG A
	vector: pQET
	Aquifex VF5 - 34LC

5' CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGATTGGC AATTTGAAAT TGA
3' CGGAGGTACC TTAAAGTGCT CTCATATCCC C KpnI

vector: pQET

Teredinibacter 42L

10

15

5

- 5' CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCCAGCT AATGACTCAC CC
- 3' CGGAAGATCT TCAACAGGCT CCAAATAATT TC (without His-tag)
- 3' CGGAAGATCT ACAGGCTCCA AATAATTTC (with His-tag)

vector: pQE12

Archaeoglobus fulgidus VC16-16MC

20

25

3" CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCTTGAT ATGCCAATCG AC

EcoR1

3' CGGAGGTACC CTAGTCGAAG ACAAGAAGAG C

Kpnl

vector: pQET

Sulfolabus solfataricus P1-8LC

5' CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCCCCAG GATCCTAGAA TT

EcoR!

3' CGGAGGTACC TTAAATTTTA TCATAAAATA C

Kpnl

vector: pQET

[0121] The restriction enzyme sites indicated correspond to the restriction enzyme sites on the bacterial expression vector indicated for the respective gene (Qiagen, Inc. Chatsworth, CA). The pQE vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori). an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites.

[0122] The pQE vector was digested with the restriction enzymes indicated. The amplified sequences were ligated into the respective pQE vector and inserted in frame with the sequence encoding for the RBS. The ligation mixture was then used to transform the *E. coli* strain M15/pREP4 (Qiagen, Inc.) by electroporation. M15/pREP4 contains multiple copies of the plasmid pREP4, which expresses the lacl repressor and also confers kanamycin resistance (Kan¹). Transformants were identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture was used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 (O.D.⁶⁰⁰) of between 0.4 and 0.6. IPTG (*IsopropyI-B-D-thiogalacto pyranoside*) was then added to a final concentration of 1 mM. IPTG induces by inactivating the lacl repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation.

50 [0123] The primer sequences set out above may also be employed to isolate the target gene from the deposited material by hybridization techniques described above.

Example 2

Isolation of a Selected Clone from the Deposited Genomic Clones

[0124] The two oligonucleotide primers corresponding to the gene of interest are used to amplify the gene from the deposited material. A polymerase chain reaction is carried out in 25 μ l of reaction mixture with 0.1 μ g of the DNA of

the gene of interest. The reaction mixture is 1.5-5 mM MgCl₂, 0.01 % (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP. 25 pmol of each primer and 1.25 Unit of Taq polymerase. Thirty cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with the Perkin-Elmer Cetus 9600 thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the gene of interest by subcloning and sequencing the DNA product.

Example 3

10 Production of the Expression Gene Bank

[0125] Colonies containing pBluescript plasmids with random inserts from the organisms M11TL, *Thermococcus* GU5L5, and *Teredinibacter* were obtained according to the method of Hay and Short, *Strategies*, *5*:16, 1992.

15 Example 4

20

30

45

Screening for Lipase/Esterase Activity

[0126] The resulting colonies were picked with sterile toothpicks and used to singly inoculate each of the wells of 96-well microtiter plates. The wells contained 250 µL of LB media with 100 µg/mL ampicillin, 80 µg/mL methicillin, and 10% v/v glycerol (LB Amp/Meth. glycerol). The cells were grown overnight at 37°C without shaking. This constituted generation of the "Source GeneBank." Each well of the Source GeneBank thus contained a stock culture of *E. coli* cells, each of which contained a pBluescript with a unique DNA insert.

[0127] The plates of the Source GeneBank were used to multiply inoculate a single plate (the "Condensed Plate") containing in each well 200 µL of LB Amp/Meth, glycerol. This step was performed using the High Density Replicating Tool (HDRT) of the Beckman Biomek with a 1 % bleach, water, isopropanol, air-dry sterilization cycle in between each inoculation. Each well of the Condensed Plate thus contained 10 to 12 different pBluescript clones from each of the source library plates. The Condensed Plate was grown for 16 hours at 37°C and then used to inoculate two white 96-well Polyfiltronics microtiter daughter plates containing in each well 250 µL of LB Amp/Meth (no glycerol). The original condensed plate was put in storage -80°C. The two condensed daughter plates were incubated at 37°C for 18 hours.

[0128] The short chain esterase '600 µM substrate stock solution' was prepared as follows: 25 mg of each of the following compounds was dissolved in the appropriate volume of DMSO to yield a 25.2 mM solution. The compounds used were 4-methylumbelliferyl proprionoate, 4-methylumbelliferyl butyrate, and 4-methylumbelliferyl heptanoate. Two hundred fifty microliters of each DMSO solution was added to ca 9 mL of 50 mM, pH 7.5 Hepes buffer which contained 0.6% of Triton X-100 and 0.6 mg per mL of dodecyl maltoside (Anatrace). The volume was taken to 10.5 mL with the above Hepes buffer to yield a slightly cloudy suspension.

[0129] The long chain '600 μ M substrate stock solution' was prepared as follows: 25 mg of each of the following compounds was dissolved in DMSO to 25.2 mM as above. The compounds used were 4-methylumbelliferyl elaidate, 4-methylumbelliferyl palmitate, 4-methylumbelliferyl oleate, and 4-methylumbelliferyl stearate. All required brief warming in a 70°C bath to achieve dissolution. Two hundred fifty microliters of each DMSO solution was added to the Hepes buffer and diluted to 10.5 mL as above. All seven umbelliferones were obtained from Sigma Chemical Co.

[0130] Fifty μ L of the long chain esterase or short chain esterase '600 μ M substrate stock solution' was added to each of the wells of a white condensed plate using the Biomek to yield a final concentration of substrate of about 100 μ M.. The fluorescence values were recorded (excitation = 326 nm, emission = 450 nm) on a plate-reading fluorometer immediately after addition of the substrate. The plate was incubated at 70°C for 60 minutes in the case of the long chain substrates, and 30 minutes at RT in the case of the short chain substrates. The fluorescence values were recorded again. The initial and final fluorescence values were compared to determine if an active clone was present.

50 Example 5

Isolation and Purification of the Active Clone

[0131] To isolate the individual clone which carried the activity, the Source GeneBank plates were thawed and the individual wells used to singly inoculate a new plate containing LB Amp/Meth. As above, the plate was incubated at 37° C to grow the cells, 50μ L of 600μ M substrate stock solution was added using the Biomek and the fluorescence was determined. Once the active well from the source plate was identified, cells from this active well were streaked on agar with LB/Amp/Meth and grown overnight at 37° C to obtain single colonies. Eight single colonies were picked with

a sterile toothpick and used to singly inoculate the wells of a 96-well microtiter plate. The wells contained 250 μ L of LB Amp/Meth. The cells were grown overnight at 37°C without shaking. A 200 μ L aliquot was removed from each well and assayed with the appropriate long or short chain substrates as above. The most active clone was identified and the remaining 50 μ L of culture was used to streak an agar plate with LB/Amp/Meth. Eight single colonies were picked, grown and assayed as above. The most active clone was used to inoculate 3 mL cultures of LB/Amp/Meth, which were grown overnight. The plasmid DNA was isolated from the cultures and utilized for sequencing.

SEQUENCE LISTING

10	[0132]
	(1) GENERAL INFORMATION:
	(i) APPLICANTS:
15	ROBERTSON, Daniel E.
	MURPHY, Dennis
20	REID, John
	MAFFIA, Anthony
oc.	LINK, Steven
25	SWANSON, Ronald V.
	WARREN, Patrick V.
30	KOSMOTKA, Anna
	CALLEN, Walter
25	(ii) TITLE OF INVENTION:
35	ESTERASES
	(iii) NUMBER OF SEQUENCES: 62
40	(iv) CORRESPONDENCE ADDRESS:
	(A) ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI, STEWART & OLSTEI
	(B) STREET: 6 BECKER FARM ROAD (C) CITY: ROSELAND
45	(D) STATE: NEW JERSEY (E) COUNTRY: USA
	(F) ZIP: 07068
50	(v) COMPUTER READABLE FORM:
50	(A) MEDIUM TYPE: 3.5 INCH DISKETTE
	(B) COMPUTER: IBM PS/2
	(C) OPERATING SYSTEM: MS-DOS (D) SOFTWARE: WORD PERFECT 5.1
<i>5</i> 5	(U) SOFTWARE, WORD FERFECT 3.1
	(vi) CLIBBENT APPLICATION DATA-

(A) APPLICATION NUMBER: Unassigned

	(B) FILING DATE: Concurrently (C) CLASSIFICATION: Unassigned	
5	(vii) PRIOR APPLICATION DATA:	
3	(A) APPLICATION NUMBER:(B) FILING DATE:(C) CLASSIFICATION:	
10	(viii) ATTORNEY/AGENT INFORMATION:	
15	(A) NAME: HERRON, CHARLES J.(B) REGISTRATION NUMBER: 28,019(C) REFERENCE/DOCKET NUMBER: 331400-39	
15	(ix) TELECOMMUNICATION INFORMATION:	
	(A) TELEPHONE: 201-994-1700 (B) TELEFAX: 201-994-1744	
20	(2) INFORMATION FOR SEQ ID NO:1:	
	(i) SEQUENCE CHARACTERISTICS	
25	(A) LENGTH: 52 NUCLEOTIDES(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: SINGLE(D) TOPOLOGY: LINEAR	
30	(ii) MOLECULE TYPE: cDNA	·
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:	
35	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGTCTTTA AACAAGCACT CT	_. 52
	(2) INFORMATION FOR SEQ ID NO:2:	
40	(i) SEQUENCE CHARACTERISTICS	
	(A) LENGTH: 31 NUCLEOTIDES(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: SINGLE(D) TOPOLOGY: LINEAR	
45	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:	
50	CGGAAGATCT CTATCGTTA GTGTATGATT T	31
	(2) INFORMATION FOR SEQ ID NO:3:	
55	(i) SEQUENCE CHARACTERISTICS	
	(A) LENGTH: 52 NUCLBOTIDES	

	(C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
5	(ii) MOLECULE TYPE: cDNA	
3	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:	
	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGAAACTC CTTGAGCCCA CA	52
10	(2) INFORMATION FOR SEQ ID NO:4:	
	(i) SEQUENCE CHARACTERISTICS	
15	(A) LENGTH: 31 NUCLEOTIDES (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
20	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:	
25	CGGAAGATCT CGCCGGTACA CCATCAGCCA C	31
	(2) INFORMATION FOR SEQ ID NO:5:	
30	(i) SEQUENCE CHARACTERISTICS	
	(A) LENGTH: 52 NUCLEOTIDES (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE	
35	(D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:	
40	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCCATAT GTTAGGAATG GT	52
	(2) INFORMATION FOR SEQ ID NO:6:	
45	(i) SEQUENCE CHARACTERISTICS	
50	(A) LENGTH: 53 NUCLEOTIDES(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: SINGLE(D) TOPOLOGY: LINEAR	
	(ii) MOLECUM: TYPE: cDNA	
55	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:	
	CGGAGGTACC TTAGAACTGT GCTGAAGAAA TAAATTCGTC CATTGCTCTA TTA	53

	(2) INFORMATION FOR SEQ ID NO:7:	
	(i) SEQUENCE CHARACTERISTICS	
5	(A) LENGTH: 49 NUCLEOTIDES (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
10	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:	·
15	CGGAGGTACC TTAGAACTGT GCTGAAGAAA TÄAATTCGTC CATTGCTCT	49
	(2) INFORMATION FOR SEQ ID NO:8:	
20	(i) SEQUENCE CHARACTERISTICS	
20	(A) LENGTH: 53 NUCLEOTIDES (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
25	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:	
30	CCGAGAATTC ATTANAGAGG AGAAATTAAC TATGAGATTG AGGAAATTTG AAG	53
	(2) INFORMATION FOR SEQ ID NO:9:	
35	(i) SEQUENCE CHARACTERISTICS	
	(A) LENGTH: 31 NUCLEOTIDES	
40	(B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	- "
40	(C) STRANDEDNESS: SINGLE	. "
40	(C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	. "
	(C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: cDNA	31
45	(C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:	31
	(C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9: CGGAGGTACC CTATTCAGAA AGTACCTCTA A	31
45	(C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9: CGGAGGTACC CTATTCAGAA AGTACCTCTA A (2) INFORMATION FOR SEQ ID NO:10:	31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGTTTAAT ATCAATGTCT TT		52
5			
	(2) INFORMATION FOR SEO ID NO:11:		
	(i) SEQUENCE CHARACTERISTICS		
10	(A) LENGTH: 31 NUCLEOTIDES.		
	(B) TYPE: NUCLEIC ACID		•
	(C) STRANDEDNESS: SINGLE		
	(D) TOPOLOGY: LINEAR		•
15	(ii) MOLECULE TYPE: cDNA	•	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:		
20	CGGAAGATCT TTAAGGATTT TCCCTGGGTA G		. 33
	(2) INFORMATION FOR SEQ ID NO:12:		
25	(i) SEQUENCE CHARACTERISTICS		
2.0	(A) LENGTH: 52 NUCLEOTIDES		
	(B) TYPE; NUCLEIC ACID		
	(C) STRANDEDNESS: SINGLE		
	(D) TOPOLOGY: LINEAR		
30	(b) for obodi. Emerin		
	(ii) MOLECULE TYPE: cDNA		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:		
35	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGGAGGTT TACAAGGCCA AA		52
	(2) INFORMATION FOR SEQ ID NO:13:		
40	(i) SEQUENCE CHARACTERISTICS		
	(A) LENGTH: 31 NUCLEOTIDES		
	(B) TYPE: NUCLEIC ACID		
	(C) STRANDEDNESS: SINGLE		
45	(D) TOPOLOGY: LINEAR		
	(ii) MOLECULE TYPE: cDNA		
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:		
	CGGAGGTACC TTATTGAGCC GAAGAGTACG A		31
55	(2) INFORMATION FOR SEQ ID NO:14:		
	(i) SEQUENCE CHARACTERISTICS		
	(A) LENGTH: 53 NUCLEOTIDES		

(B) TYPE: NÚCLEIC ACID

	(C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
5	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
10	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGATTGGC AATTTGAAAT TGA	53
	(2) INFORMATION FOR SEQ ID NO:15:	
15	(i) SEQUENCE CHARACTERISTICS	
	(A) LENGTH: 31 NUCLEOTIDES(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: SINGLE(D) TOPOLOGY: LINEAR	
20	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:	
25	CGGAGGTACC TTAAAGTGCT CTCATATCCC C	31
	(2) INFORMATION FOR SEQ ID NO:16:	
30	(i) SEQUENCE CHARACTERISTICS	
35	(A) LENGTH: 31 NUCLEOTIDES (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: cDNA	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:	
	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCCAGCT AATGACTCAC CC	52
45	(2) INFORMATION FOR SEQ ID NO:17:	
45	(i) SEQUENCE CHARACTERISTICS	
50	(A) LENGTH: 32 NUCLEOTIDES (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: cDNA	
55	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:	٠
	CGGAAGATCT TCAACAGGCT CCAAATAATT TC	32

	(2) INFORMATION FOR SEQ ID NO:18:		
	(i) SEQUENCE CHARACTERISTICS		
5	(A) LENGTH: 29 NUCLEOTIDES (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR		
10	(ii) MOLECULE TYPE: cDNA		•
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:		
15	CGGAAGATCT ACAGGCTCCA AATAATTTC		29
	(2) INFORMATION FOR SEQ ID NO:19:		
20	(i) SEQUENCE CHARACTERISTICS		
20	(A) LENGTH: 52 NUCLEOTIDES (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR		
25	(ii) MOLECULE TYPE : cDNA		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:		
30	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCTTGAT ATGCCAATCG AC		52
35	(2) INFORMATION FOR SEQ ID NO:20:		
	(i) SEQUENCE CHARACTERISTICS		
40	(A) LENGTH: 31 NUCLEOTIDES (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR		,
	(ii) MOLECULE TYPE: cDNA	•	
45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:		
	CGGAGGTACC CTAGTCGAAC AGAAGAAGAG C		31
50	(2) INFORMATION FOR SEQ ID NO:21:		
	(i) SEQUENCE CHARACTERISTICS		
55	(A) LENGTH: 52 NUCLEOTIDES(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: SINGLE(D) TOPOLOGY: LINEAR		

	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:	
5	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCCCCTA GATCCTAGAA TT	52
	(2) INFORMATION FOR SEQ ID NO:22:	
10	(i) SEQUENCE CHARACTERISTICS	
15	(A) LENGTH: 31 NUCLEOTIDES(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: SINGLE(D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: cDNA	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
	CGGAGGTACC TTAAATTTTA TCATAAAATA C	31
25	(2) INFORMATION FOR SEO ID NO:23:	
	(i) SEQUENCE CHARACTERISTICS	
30	(A) LENGTH: 555 NUCLEOTIDES(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: SINGLE(D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: GENOMIC DNA	·
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
40		

		Sex									Met					Leu	46
5					Pro					Ala					Met	TCG Ser	96
10									Ala							GCT	144
	GTT Val	AAA Lys 50	Ser	AGT Ser	GTT Val	GCT Ala	CCA Pro 55	ATA Ile	TTC Pbe	AAT Asn	CTA Leu	GGT Gly 60	CTT	GCA Ala	ATT	Gly	192
15	GGG Gly 65	ATA Ile	CTA Leu	ATT Ile	GIT Val	ATA Ile 70	GIT Val	GGT Gly	TTA Leu	AGA Arg	AAT Asn 75	CTT	TAT Tyr	TCG Ser	Teg	AGT Ser 80	240
20																TTA Leu	288
	ATA Ile	G17 GGG	GTT Val	TTC Phe 100	GAC Asp	GAA Glu	GTA Val	TAT Tyr	GGT Gly 105	TGG Txp	ATA Ile	CAT His	TTC Phe	CTA Leu 110	GTC Val	TCA Ser	336
25			TTT Phe 115													ATA Ile	384
			AAA Lys								Ile						432
30			TAT Tyr														480
35	,	·~ ~	AG T	DA TE	ra <i>cc</i>	יני כד	·	C TC	ः गन् न	r ques			71 7	ATTA	202	C10	
			lu L			a Va					Pro			Ile			528
40			TT AU he Ly		r Ty					;							576
45	(2)	INFOF	RMATI	ON FO	OR SE	Q ID I	NO:24	:									
		•	QUEN														
50		(I (I	A) LEN B) TYF C) STF D) TOI	PE: NU	JCLEI EDNE	C ACI SS: SI	D INGLE										
		(ii) M	OLEC	ULE T	YPE:	GENC	MIC E	DNA									
55		(xi) S	EQUE	NCE I	DESC	RIPTIO	ON: SI	EQ ID	NO:2	4:							

											r Ty					GAT ASP	
5	TTA		TIG	CAI	TII	GCA	TT	TAC	TG	T	cro	GCC	GTC	TAT	r acc	TGG	
	-	_	Leu	His 20		Ala	Phe	Tyz	Trp 25		e Lev	ı Ala	Va]	.Ty:		Trp	
10	TTA Leu	Pro	GGT Gly 35	GTC Val	CTA Leu	GTC Val	CGG Arg	GGC Gly 40	GTA Val	GCI	GIG Val	GAC Asp	ACA Thx 45	GGG	GTG Val	GCT Ala	144
	CGG Arg	GTG Val 50	Pro	GCG	CTC Leu	GGC	CGG Arg 55	CGC	GGT	Lys Lys	AGG Arg	CTG Leu 60	CTC	CTG Leu	GCC	GCT	192
15			GTC Val									Val				GTG Val 80	240
20	GCG Ala	TAT Tyr	AGT Ser	AGT Ser	CTG Leu 85	CAC His	CCG Pro	GAG Glu	AGC Ser	TGT Cys 90	Arg	Pro	GTT Val	GCG Ala	CCG Pro 95	GAG Glu	288
	GGG · Gly															GTG Val	336
25	GTT Val		GGC Gly 115														384
	GTT Val																432
30	GTG Val 145														Val		480
35	GAC :																528

					165	· .				170)				175	•	-	
5	CGG	GAG Glu	GTG Val	Leu 180	qeA ı	GCC	CGG Arg	GCI Ala	Val	Val	Gly	TAT	Val	Ser 19	Glu	CGG Arg		576
٠.	TTC Phe	Pro	GGC Gly 195	Arg	CGG Arg	ATA	ATA	Leu 200	Val	Gly	TIC Phe	AGT	Met 205	Gly	Gly	GCT Ala		. 624
10	GTA Val	GCG Ala 210	ATC Ile	GTG Val	GAG Glu	GGT Gly	GCT Ala 215	GGG	GAC Asp	CCG Pro	CGG	GTC Val 220	TAC Tyr	GCG Ala	GTG Val	GCT Ala		672
15	GCT Ala 225	gat Asp	AGC Ser	CCG	TAC Tyr	TAT Tyr 230	agg Arg	CTC	CGG	A ap GAC	GTC Val 235	ATA Ile	CCC	CGG Arg	TGG Trp	CTG Leu 240		720
	GAG Glu	TAC Tyr	aag Lys	ACG Thr	CCG Pro 245	CTG Leu	CCG Pro	G17 GCC	TGG Trp	GIG Val 250	GIY Gly	GIG Val	CIG Leu	GCC Ala	GGG Gly 255	TTC Phe		768
20	TAC Tyr	GGG Gly	AGG Arg	CTG Leu 260	ATG Met	GCG Ala	GJ y GGC	GTT Val	GAC Asp 265	CTC Leu	GCC	TIC Phe	GJY	CCC Pro 270	GCT Ala	GGG		816
25					GAT Asp													864
					CGG Arg													912
30	TGT Cys 305	GGC	CGT Arg	CTC Leu	GTC Val	GAG Glu 310	GTI Val	CCT Pro	GGG Gly	GCT Ala	GGC Gly 315	CAC His	GTG Val	GAG Glu	GCC Ala	GTG Val 320		960
35	gat Asp	GTG Val	CTC Leu	GGG GGG	Pro 325	Gly	CGC Arg	TAC Tyr	GCA Ala	GAC Asp 330	ATG Met	CTG Leu	ATA Ile	Glu	CIG Leu 335	GCG Ala		1008
ω	CAC His		Glu		CCT Pro						TGA					-		1019
40	(2) II	NFOF	IMATI	ON F	OR SE	:O ID I	NO:25	:										٠

(2) INFORMATION FOR SEO ID NO:25:

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 789 NUCLEOTIDES
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: GENOMIC DNA

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

5	ATG CC Met Pr 1	A TAT	GTT AGG	AAT (GGT GG 31y G1	T GTA y Val	AAT ASD 1	NTC TA	TAT TYP	GAA Glu	CTG Leu 15	GTG Val	48
	GAT GG Asp Gl	A CCT y Pro	GAG CCI Glu Pro 20	CCA /	ATT GT Tle Va	C TTT 1 Phe 25	Val H	ia Gl	A TGG y Trp	ACA Thr 30	GCA Ala	AAT Asn	96
10	ATG AA Met As	r TTT n Phe 35	TGG AAA TYP Lys	GAG (AA AG In Arg	g Arg	TAT T	TT GC he Al	A GGC a Gly 45	AGG Arg	AAT Asn	ATG Met	144
15	ATG TTO Met Les 50	ı Phe '	GTC GAT Val Asp	AAC A Asn A	GA GGT LTG Gly 55	CAT His	GGC A	GG TCC rg Sea 60	dey 3	AAG Lyb	CCA Pro	CTT Leu	192
	GGA TAG Gly Tyr 65	GAT 1 Asp 1	Phe Tyr	AGA T Arg P	TT GAG be Glu	AAC Asn	Phe I	TT TC: le Ser 75	GAT Asp	TTA (Aap 1	CG Vla 00	240
20 [.]	GTT GTT Val Val	AGG G	AG ACT llu Thr 85	GGA G Gly V	TG GAG al Glu	Lys	TTT GI Phe Va 90	T CTC	orc Val	GGA (Gly E	CAT 1 lis s 95	CA. Ser	288
,	TTC GGA Phe Gly	Thr M	et Ile	Ser Me	et Lýs	Tyr (Cys Se	r Glu	Tyr	Arg A 110	A COR	rg	336
25	GIT CIT Val Leu	Ala L 115	eu Ile	Leu Il	le Gly 120	Gly C	Sly Se	r Arg	Ile 1 125	Lys L	eu L	eu	384
30	CAC AGA His Arg 130	Ile G	ly Tyr	Pro Le 13	u Ala	Lys I	le Le	Ala 140	Ser 1	le A	la T	γ ι	432
	AAG AAG Lys Lys 145	Ser Se	er Arg	Leu Va 150	l Ala	Asp L	eu Sei 15!	r Phe	Gly I	ув А	92 A.	la 50	480
35	GGT GAA Gly Glu TCC TAC	Leu Ly	ys Glu 165	Trp Gl	y Trp	Lya G	ln Ala 70	Met	Asp T	yr 71 13	nr Pr 75	r o	528
	Ser Tyr	Val Al	la Met 1 30	lyr Th	r Tyr	Arg T 185	hr Lev	Thr	Lys V	al As 190	n Le	eu	576 624
40	Glu Asn GAA GAG	Ile Le 195	eu Glu	Lys Il	e Asp 200	Cys P	ro Thr	Leu	Ile I 205	le Va	l Gl		672
45	Glu Glu 210 AGG ATA	Asp Al	la Leu I	eu Pro 21	o Val . 5	Ser Ly	ys Ser	Val 220	Glu L	eu Se	r Ar	9	720
	Arg Ile 225	Glu As	n Ser I	ys Lei !30	u Val	Ile Il	le Pro 235	Asn :	Ser G	ly Hi	в Су 24	0	768
50	Val Met						m Arg				u Ph		

ATT TCT TCA GCA CAG TTC TAA Ile Ser Ser Ala Gln Phe 260

55

⁽²⁾ INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTER

- (A) LENGTH: 756 NUCLEOTIDES
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULAR TYPE: GENOMIC DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

15									Let					d GGA		48
				Ile				Val					AST	GAG Glu		96
20			Arg									Gly		ATC		144
25		Val									Gly			AGC Ser		192
25														Pro 80	٠.	240
30														GAA Glu .		288
										ATA Ile				ATA Ile		336
35 ·										TAC Tyr				Gly		384
40	Ser									ATT Ile 140						432
										gac Asp						480
45			Pro				Gln			GCT Gly		Pro				528
		qeA				Glu				GAT . QaA			Asn			576

55

50

	C1 Le	T (CAC His	116 195	Le	T TT u Le	G AG u Ar	g ag	C TT Ph	e Ph	T CT e Le	T GC	G GI a Va	C CG 1 Ar 20	g Se	A AA E AB	C TCC n Ser	:	624
5 ·	GA G1	u I	AG yb 10	AG/ Arg	AA Ly:	G GA	G TT u Pb	T TG: e Cyr 215	a yel	C CTO	C GT u Va	T AT.	A GT e Va 22	l Pr	T GA o G1	G CT u Le	T GAG u Glu	3	672
10	GA G1 22	u P	TC	ACA Thr	. Pro	C CT	CA: 1 Asj 230	o Val	AG/	A AAI J Lys	A GCC	GA(LAB) 23!	Gl	A AT	A AT	G GAI	AGG Arg 240		720
							Let	A GAG				Gli		3	•				768
15	(2) IN	IFOI	RMA	TIOI	V FO	R SEC	A OI C	10:27:											
	(i) SE	QU	ENC	E CH	ARA	CTER	ISTIC	S										
20		(B) T C) S	YPE	: NU(DNES	ACIE	NGLE											
25								OMIC ON: SE		NO:2	27:								
30	ATG Met	Pl	rr i ne i	TAA Asn	ATC Ile	AAT Ass 5	Val	TIT Phe	GIT Val	AAT Asn	TATA	Sex	TI	CTC Lev	TA:	f TTT	TCA Ser		48
											Tyr					Thr	GC		96
35															Phe		ACT		144
40		11						GGA Gly 55									ATT		192
	AGT Ser 65	GT Va	T G	CT Lla	GAA Glu	GAA Glu	TTT Phe 70	GCT Ala	AGG Arg	CAC His	GGA Gly	TTT Phe 75	GGA Gly	TTC	TGC Cys	ATG Met	CAC His 80		· 240
45								AGA Arg											288
				ly				TTC Phe											336
50			a L					GGA Gly											384
55			E G				Ile	GCG Ala 135											432
	ATC	GCC	: A	AG (GA ·	GIT	ATC	GCG	CTA	GCC	cca	GCC	CTC	CAA	ATC	ccc	TTA		480

	I1 14		a Lye	Gl _y	va.	150		Let	ı Ala	a Pro	155		ı Glı	n Il	e Pro	160	
5	AC:	CCG PEC	GCT Ala	AGA	AGA Arg 165	Lev	GII Val	CTA Leu	AGC Ser	Lev 170	ı Ala	TCA Sex	AGG Arg	CT Let	GCC 1 Ala 175	CCG Pro	528
	CA:	r TCI Ser	Lys	ATC Ile 180	Thr	TTA Leu	CAA Gln	AGG Arg	AGA Arg 185	Leu	Pro	CAG Gln	Lys	Pro	Glu	GGT	576
10	TTT Phe	CAA Gln	AGA Arg 195	GCA Ala	aaa Lys	CAT QBA	ATA Ile	GAA Glu 200	TAC Tyr	AGT Ser	CTG Leu	AGT Ser	GAA Glu 205	ATA Ile	TCA Ser	GIC Val	624
15	AAG Lys	CTC Leu 210	Val	GAC Asp	GAA Glu	ATG Met	ATT Ile 215	AAA Lys	GCA Ala	TCA Ser	TCT Ser	ATG Met 220	TTC Phe	Trp Trp	ACC	ATA Ile	672
	GCA Ala 225	Gly	GAA Glu	ATT Ile	AAT ABD	ACT Thr 230	Pro	GTC Val	CTG Leu	CIT Leu	ATT Ile 235	CAT Ris	GGG Gly	GAA Glu	aaa Lyb	GAC Asp 240	720
20	TAA Asa	GTC Val	ATA Ile	CCT Pro	CCG Pro 245	GAG Glu	GCG Ala	AGC Ser	AAA Lys	AAA Lys 250	GCC	TAC Tyr	Gln Gln	Leu	A ATI Ile 255	Pro	768
25			Pro														816
	Phe	TIT Phe	GAA Glu 275	CCA Pro	G1y GGC	GCG Ala	Val	AAA Lys 280	ATC Ile	GTC. Val	ACA Thr	Asp	ATT Ile 285	GTA Val	GAG Glu	TCG Trp	864
30			TAA mea		_					TAA							874
	(2) !!	NFOR	MATIC	N FO	R SE	QĮDI	VO:28	:									
35	((i) SEC	QUEN	CE CH	HARA	CTER	ISTIC	s									
40		(B (C) LENG) TYPI) STR) TOP	E: NU ANDE	CLE	C ACII SS: SI	D NGLE	_									
	. ((ii) MO	LECU	LE T	/PE: (GENO	MIC E	ONA									
45	((xi) SE	QUEN	NCE D	ESCI	RIPTIO	ON: SI	EQ ID	NO:2	28:							
•																GTC Val	48
50	GIT Val	CTG Leu	GTT (Val	CAT His 20	GGC Gly	CTC Leu	Gly	GAG Glu	CAC His 25	AGC Ser	GGA Gly	AGG Arg	TAT Tyr	GGA Gly 30	AGA Arg	CTG Leu	96
55	ATT Ile	aag Lyb	GAA (Glu) 35	CTC . Leu .	AAC Asn	TAT Tyr	GCC Ala	GGC Gly 40	TTT Phe	GGA Gly	GTT Val	TAC Tyr	ACC Thr 45	TTC Phe	GAC Asp	TGG Trp	144
	ccc	GGC	CAC (GGG .	AAG	AGC	CCG	GGC	AAG	AGA	GGG	CAC	ACG	AGC	GTC	GAG	192

		Pro	5 G1	0 A H7	a GI	у Гу	8 Se:	r Pro	o G1 5	y Ly	B Ar	g Gl	y Hi	s Th O	ur Se	r Va	l Gl	u	
<i>5</i>		GAG Glu 65	Ali	g at a Me	G GA t Gl	A ATG	C ATO	e Aej	C TC P Se	G AT	A AT	C GAI e Gli 75	Gl:	S AT	C AG e An	G GA g Gl	G AA U Ly B	9	240
		Pro	TIC Phe	CTO Let	TTO Pho	C GG(e Gl) 85	, Hie	AGC Sex	CT Lei	c GGT	GG; Gly 90	Y Let	A ACT	Val	C AT	C AG Ar	G TAI g Ty: 5	C F	288
10						Pro					Gly					: Se	G CCT		336
15		GCC Ala	CTC Leu	GCC Ala 115	Lys	AGC Ser	CCG Pro	GAA Glu	ACG Thr 120	Pro	Gly	Phe	ATG Met	Val 125	Ala	Lev	GCG Ala	;	384
		aag Lys	TTC Phe 130	CTT Leu	GGA Gly	AAG Lys	ATC Ile	GCC Ala 135	CCG	GGA Gly	GIT Val	Val	CTC Leu 140	TCC Ser	AAC	GGC Gly	ATA Ile		432
20	1	aag Lys 145	CCG Pro	GAA Glu	CTC Leu	CTC Leu	TCG Ser 150	AGG Arg	AAC Asn	AGG Arg	GAC Asp	GCC Ala 155	GTG Val	AGG Arg	AGG Arg	TAC	GIT Val 160		480
	6	BAA Blu	QAC Asp	CCA Pro	CTC Leu	GTC Val 165	CAC His	GAC Asp	AGG Arg	ATT Ile	TCG Ser 170	GCC Ala	aag Lys	crg Leu	gga Gly	AGG Arg 175	AGC Ser		528
25										CAC His 185						Ile			576
30								Ile		ACT Thr			Val						624
		Slu (Gly 210	Ser	Arg	Arg	Leu	Phe 215	Glu	GAG Glu	Leu	Ala	Val 220	Glu	Asd	Lув	Thr		672
35	r· 1	leu 1 25	Arg	Glu	Phe	Glu	Gly 230	Ala	Tyr	CAC His	Glu	11e 235	Phe ·	Glu .	qaa	Pro	Glu 240		720
40	7	ן פבני	Ala	Glu	Glu	Phe 245	His (Glu '		ATT Ile					Val (768
40				Ser		GCT (Ala (CAA ' Gln	TAA											775
45	• • •						ID NO												
		•	•			0 NU0 EIC A	CLEO.	TIDES	6										
50		(C) S	TRAN	DEDI		: SINC	SLE											
		(ii) M	OLE	CULE	TYP	E: GE	NOMI	C DN	A										
55		(xi) S	EQU	ENC	E DES	SCRIP	NOIT	: SEC) ID N	O:29:									

÷	Le				n Lei						e Gl				n Le	ra GMT u Val .5	
5					y Ala					e Ala						a AAA u Lys	
10	GC.	CRC Let	GAJ G1t 35	ı Glı	CTC	GGT Gly	ATA Ile	AA(Lyi	ya]	A AAG L Lys	AGC Arg	G Leu	2 AG 1 Se 4	C1	G GT y Va	A AGT 1 Ser	144
	GCT Ala	ı Gly	A GCI y Ala	T ATO	GTT Val	TCC	Val	Phe	TAC Tyr	GCI Ala	TCC Ser	60 60	TAC Ty:	AC.	r CC	C GAC	192
15		Met					Lys					Leu				DAA 1 Byl 9 08	240
20						Met					Trp					A GAG a Glu ;	288
					Glu					Arg					Ası	ATA Ile	336
25				Leu					Leu					Ala		TAC	384
			Arg													ATA Ile	432
30	CCC Pro 145	Gly	ATT	TTT	GAA Glu	CCA Pro 150	GIT Val	GAG Glu	TAC Tyr	GAG Glu	AAT Asn 155	TTT Phe	CTA Leu	CIT	GIT Val	GAC Asp 160	480
35	Gly	Gly	Ile	Val	Asn 165	Asn	Leu	Pro	Val	Glu 170	Pro	Leu	Glu	Lys	Phe 175		528
	GAA Glu	CCC Pro	ATA Ile	ATC Ile 180	GGG Gly	GTA Val	gat Asp	GTG Val	CTT Leu 185	CCC Pro	ATA Ile	ACT Thr	CAA Glb	GAA Glu 190	Arg	aag Lys	576 `
40	ATT Ile	AAA Lye	AAT Asn 195	ATA Ile	CTC Leu	CAC . His	Ile	CTT Leu 200	ATA Ile	AGG Arg	AGC Ser	TTC Phe	TTT Phe 205	CTG Leu	GCG Ala	GTT Val	624
					GAA Glu	Lys .					Сув						672
45					GAG Glu					Asp							720
50																	

(2) INFORMATION FOR SEQ ID NO:30:

55

(i) SEQUENCE CHARACTERISTICS

ATA TTC TGC GGG GAT ATG AGA GCA CTT TAA Ile Phe Cys Gly Asp Met Arg Ala Leu 245

- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

50

55

(ii) MOLECULE TYPE: GENOMIC DNA

(xi) SEQUENCE DESCRIPTION: SEQ In NO:30:

10																		
		Pro									Phe					: ATT		48
15					His					Leu					Leu	CGC Arg		96
									Lys					Thr		CAA Gln		144
<i>2</i> 0			Glu										Ala			CTT Leu		192
25		Thr									CAC His 75					AAC Asn 80	240	
											GGC Gly							288
30											GAC Asp						•	336
											ACC Thr							384
35											GTA Val							432
40											TAT Tyr 155							480
										_	GCG Ala			_	_	_		528
45		Ala					Gly				GTA Val	Cys						576
	ccc	GCA	CAC	ACC	ATG .	ATG (GCC	CTA .	AAC	CGA	GGT (GCG	TIT	TTC	TAC	GGC		624

•	Pro	195	His	Thi	. Met	: Met	200	Lev	l Agr	Ar	g G1 ₃	7 Ala 205		e Ph	e Ty	r Gly	
5	CGC Arg 210	Tyr	Phe	GCG Ala	CAT His	Lys 215	Trp	Lys	Arg	Ser	220	t Thr	GC/ Ala	A AAI a Lys	A CT.	GCA Ala 225	67;
	GCT Ala	TIC	CCA Pro	GAC Asp	TAC Tyr 230	Lys	TAC	GGC	AAA Lys	GAT Asp 235	Leu	Lys	Ser	ATA Ile	CAC His 240	ACG	720
10	CTT	GAT Asp	GAG Glu	TTA Leu 245	AAC Asn	AAC Asn	TAT Tyr	TTC Phe	ATT Ile 250	Pro	Arg	TAC	ACC	GGC Gly 255	Phe	AAC Asn	768
15														Gln		CIC Leu	916
	GCG Ala	TTT Phe 275	CTC Leu	AAC Asn	TGC Cys	Pro	AGT Ser 280	TAC Tyr	ATT Ile	CTG Leu	GCA Ala	GCT Ala 285	GGC Gly	QAC Asp	GAC Asp	CCA Pro	864
20						GAC Asp 295											912
25				Val		CAA Gln			Ser								960
			Lys			GCT Ala		Asp					Lys				1,008
30	GCC Ala		TGA														1,111

- (2) INFORMATION FOR SEQ ID NO:31:
- (i) SEQUENCE CHARACTERISTICS

35

40

45

- (A) LENGTH: 936 NUCLEOTIDES
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: GENOMIC DNA.
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

		,					-				
									GAG Glu 15		48
		-		_							
•	.5								GAG Glu		96
	10								CTG Leu		44
									GCG Gly	1	.92
	15										

	P.	ic GC m G1 5	iy A	sp II	le Ar	y Va	l Ar	A GT g Va	T TA 1 Ty	C CA T G1	G CA B G1: 7:	n Ly	E P	C G C A	AT T	er	Pro 80		240
5	GG Va	T CI	NG GT	T TA il Ty	C TA T Ty 8	T CA T Hi 5	C GG B Gly	r GG	r co	A TT y Pho 90	e Val	G AT	T TO e Cy	E AC	r I	TC le 95	gag Glu		288
10	Se	G CA r Hi	C GA .s A.s	C GC p Al 10	a Le	A TG	C AGO S Arg	AGI ARI	A AY: 3 Ile 105	e Ala	AGA Arg	CT:	T TC u Se	A AA r As 11	n Se	er er	ACC Thr	•	336
				r Va		TAC Tyr			ı Ala					s Ph				•	384 .
15			l Ty			TAC Tyr		Ala					. Ale					٠	432
20	Glu 145	Glu	Let	a Arg	; Ile	GAC Asp 150	Pro	Ser	Lys	Ile	Phe 155	Val	Gly	Gly	AB	p 9	ex 60		480
	Ala	Gly	Gly	' Asn	165	GCC Ala	Ala	Ala	Val	Ser 170	Ile	Met	Ala	Arg	17!	S	er	•••	`528
25 .	Gly	Glu	Asp	Phe 180	Ile	AAG Lys	His	Gln	11e 185	Leu	Ile	Tyr	Pro	Val 190	۷al	A	9n		576
. •	Phe	Val	Ala 195	Pro	Thr	CCA Pro	Ser	Leu 200	Leu	Glu	Phe	Gly	Glu 205	Gly	Leu	T :	LD.		.624
30	Ile	Leu 210	Asp	Gla	Lys	ATA Ile	Met 215	Ser	Trp	Pbe	Ser	Glu 230	Gln	Tyr	Phe	Se	er		672
35	Arg 235	Glu	Glu	Aso	Lys	TTC Phe 240	Asn	Pro	Leu	Ala	Ser 245	Val	Ile	Phe	λla	25	P		720
	Leu	Glu	Asn	Leu	Pro 255	Pro	Ala :	Leu	Ile	Ile 260	Thr I	ala.	Glu	Tyr	Авр 265	Px	0		768
40	Leu	Arg	Ąsp	G1u 270	Gly	GAA Glu	Vaļ :	Phe (Gly (275	Glm 1	Met 1	Leu ;	Arg	Arg 280	Ala	G1	y		816
·	Val	Glu	Ala 285	Ser	Ile	GTC :	Arg :	191) 290	Arg (31y 1	/al I	Leu I	H18 295	GJA	Phe	11	е		864
45	Asn	300	Tyr	Pro	Val :		Lys # 105	Ua A			k qa								912
50	GCC (Ala / 315				Val			AG											936

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 918 NUCLEOTIDES (B) TYPE: NUCLEIC ACID

- (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR

55

- (ii) MOLECULE TYPE: GENOMIC DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

10	AT(Me)	s cc t Pr	C CI	A GA! 1 ABJ	Pro	r AGA	A ATT	r aai E Ly:	A AA	G TT S Le	u Le	A GA u G1	A TC u Se	A GC r Al	a Le	TT A	or Thr		48
	AT/	A CC	A ATT	GG Gly 20	, Lys	A GCC	CCA Pro	Val	CAL COL CL CL CL CAL CL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CL CAL CA	ı Gl	G GI	A AG 1 Ar	a aa g Ly	s Il	A TT e Ph 0	T A 18 A	ra GG		96
15	CA:	TT:	A GCC a Ala 35	Sex	GCA Ala	GCI Ala	Pro	Lys 40	Va.	GAU Glu	A GT.	r GGI	A AAI y Ly: 4!	ya.	A GA 1 G1	A G	AT sp		144
20 .	ATA	Lys 50	ATA Ile	CCA Pro	GGC Gly	AGT Ser	GAA Glu 55	Thr	GTI Val	ATZ Ile	AAC : Asr	GCT Ala 60	Arg	Va.	G-TA	T T	rr he		192
	Pro 65	Lys	AGT Ser	Ser	Gly	Pro 70	Tyr	Gly	Val	Leu	.75	Tyr	Leu	Hi:	G Gl	G]	y IO		240
25	Gly	Phe	GTA Val	Ile	G1y 85	Asp	Val	Glu	Ser	77. 90	qaA	Pro	Leu	Cys	Arg 95	Al	a		288 -
30	Ile	Thr	AAT	Ala 100	Сув	Asn	Сув	Val	Val 105	Val	Ser	Val	Авр	Tyr 110	Arg	Le	u		336
	Ala	Pro	GAA Glu 115	Tyr	Lys	Phe	Pro	Ser 120	Ala	Val	Ile	Ąsp	Ser 125	Phe	Asp	Al:	a		384
35	Thr	ASD 130	TEG	Val	Tyr	Asn .	Asn 135	Leu	qaA	Lys	Phe	Asp 140	Gly	Lys	Met	Gl	Ÿ		432
	Val 145	Ala	ATT	Ala	Gly	Asp 150	Ser .	Ale	Gly	Gly	Asn 155	Leu	Ala	Ala	Val	Va]	l)		480
40	Ala .	Leu	CTT	Ser	Lys 165	Gly 1	Lys :	Ile /	Asn	Leu 170	Lys	Тух	Gln	Ile	Leu 175	Va]	Į.	•	528
45	Tyr	Pro		Val 180	Ser :	Leu i	yab 1	len '	Val . 185	Ser .	Arg	Ser :	Met	11e 190	Glu	Tyr	•		576
	Ser i	yab	GGG Gly 195	Phe	Phe	Leu 1	Chr J	urg (SAG (Slu)	His :	ATA (Glu 1	TGG Trp 1 205	ITC Phe	GIY	TCT Ser	•	(624

	CAĀ	TAC	TTA	CGA	AGC	CCI	GCA	GAT	TTG	CTA	GAC	TTT	AGG	TTC	TCI	CCA		672
	GLn	210		Arg	ser	Pro	215	Asp	Leu	Leu	Asp	220		Phe	. Sex	Pro		
; ·																ACA Thr		720
	225			44.	nop	230		GIJ	Deu	710	235	AIG	D EU	416	116	240		
•																AAA		768
o	Ala	Glu	Tyr	Asp	Pro 245	Leu	Arg	Asp	Gln	Gly 250	Glu	Ala	Tyr	Ala	Asn 255	Lys		
						GTC												816
	Leu	Leu	Gin	260	gtÀ	Val	ser	Val	7NT 265	ser	Val	Arg	Pne	270	Asn	Val		
_						TCA												864
5	Ile	His	Gly 275	Phe	Leu	Ser	Phe	Phe 280	Pro	Leu	Met	Glu	Gln 285	Gly	Arg	ДВР		
						GGG											-	912-
	Ala	11e 290	Gly	Leu	Ile	Gly	Ser 295	Val	Leu	Arg	Arg	Val 300	Phe	Tyr	Asp	Lys		
0	ATT	TAA																918
	Ile 305																	

(2) INFORMATION FOR SEQ ID NO:33:

25

30

40

45

50

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 184 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met 1	Ser	Leu	Asn	Lys 5	His	Ser	Trp	Met	Asp 10	Met	Ile	Ile	Phe	Ile 15	Leu
Ser	Phe	Ser	Phe 20	Pro	Leu	Thr	Met	Ile 25	Ala	Leu	Ala	Ile	Ser 30	Met	Ser
Ser	Trp	Phe 35	Asn	Ile	Trp	Asn	Asn 40	Ala	Leu	Ser	Авр	Leu 45	Gly	His	Ala
Val	Lys 50	Ser	Ser	Val	Ala	Pro 55	Ile	Phe	Asn	Leu	Gly 60	Leu	Ala	Ile	Gly
Gly 65		Leu	Ile	Val	Ile 70	Val	Gly	Leu	Arg	Asn 75	Leu	Tyr	Ser	Trp	Ser 80
Arg	Val	Lys	Gly	Ser 85	Leu	lle	lle	Ser	Met 90	GJÀ	Val	Phe	Leu	Asn 95	Leu
lle	Gly	Val	Phe 100	qeA	Glu	Val	Tyr	Gly 105	Trp	Ile	His	Phe	Leu 110	Val	Ser
Val	Leu	Phe 115	Phe	Leu	Ser	Ile	Ile 120	Ala	Tyr	Phe	Ile	Ala 125	Ile	Ser	Ile
Leu	Asp 130	Lys	Ser	Trp	Ile	Ala 135	Val	Leu	Leu	Ile	Ile 140	Gly	His	Ile	Ala

Met Trp Tyr Leu His Phe Ala Ser Glu Ile Pro Arg Gly Ala Ala Ile 145

Pro Glu Leu Leu Ala Val Phe Ser Pho Leu Pro Phe Tyr Ile Arg Asp 165

Tyr Phe Lys Ser Tyr Thr Lys Arg

(2) INFORMATION FOR SEQ ID NO:34:

15

20

25

30

35

40

45

50

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 346 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

	Met	Ly L	rs L	eu I	eu (Glu 5	Pro	Th	r As	n T	br S	er 10	Tyr	Th	r Le	u Le		n As .5	P
5	Lev	a Al	a L	eu H	is : 20	Phe	Ala	Pho	е Ту		rp P 25	be :	Leu	Ala	a Va	1 TY 3		r Tr	P
	Lev	Pr		ly V 35	al 1	Leu	Val	Arg	g G1 4	y Va O	al A	la '	Val	Asī	4:	r Gl	y Va	l Al	a
10	Arg	Va 5		ro G	ly 1	Leu	Gly	Arg 55	J Ar	g G	ly L	ys I	Arg	Leu 60		u Lei	u Al	a Ala	A
	Val 65		a Va	1 L	eu /	Ma	Leu 70		l Va	l Se	r V	al \	/al 75	Val	Pro	o Ala	a Ty	r Va.	
15	Ala	Ty	r Se	r S	er 1	85	His	Pro	Gl1	u Se		ув A 90	urg	Pro	Va]	l Ala	9!	-	a
•				10	00					10	5					Gly 110)	•	
20			11	5					120	•	-				125				
		130)					135						140		Тух			
25	145						150					1	55			Val		160	1
					1	65					17	0				Ile	175		
30				18	0					18	5		_			Ser 19	0		
			19	5					200	+					205	Gly			
35		210)					215					2	220		Ala			
	225	ASP	se:	r PI	O 17		230	Arg	Leu	AŊ	g as		35 31	rie	Pro	Arg	Trp	Leu 240	
40	d	31ų	Tyr	Lys	Th	r P: 24	ro 1	Leu	Pro	Gly	Tx		al (50	31y	Val	Leu	Ala	Gly 255	Phe
	1	Ŋτ	Gly	Arg	Le 1	u Mo O	et A	lla (Gly	Val	26.	р <i>L</i> e 5	eu (3ly	Phe	Gly	Pro 270	Ala	Gly
45	V	al	Glu	Arg 275	Va:	l As	sp L	ys i	Pro	Leu 280	Lev	ı Va	ıl V	/al	Tyr	Gly 285	Pro	Arg	Авр
	P	ro	Leu 290	Val	Thi	: Ar	g A	sp (31u 295	Ala	Arg	, Se	r L	eu .	Ala 300	Ser	Arg	Ser	Pro
50		ys (05	Sly	Arg	Lev	Va		lu 1 10	/al	Pro	Gly	Al		ly 1 15	His	Val	Glu	Ala	Val 320
	A	sp 1	/al	Leu	Gly	Pr 32		ly A	rg :	lyr	Ala	A81		et 1	Leu	Ile	Glu	Leu 335	Ala
55	H	ls G	lu (Glu	Сув 340	Pr	o Pi	co G	ly 3	lla	Gly 345	Gly	•						

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 262 AMINO ACIDS
 - (B) TYPE: AMINO ACID

5

10

15

20

25

30

35

40 .

50

55

- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met 1		Tyr	Val	Arg 5	Asn	Gly	Gly	Val	Asn 10		Tyr	Tyr	Glu	Leu 15	Val
Asp	Gly	Pro	Glu 20	Pro	Pro	Ile	Val	Phe 25		His	gly	Trp	Thr 30	Ala	Asn
Met	. Asn	Phe 35		Lys	Glu	Gln	Arg 40	_	Tyr	Phe	Ala	Gly 45	Arg	Asn	Met
Met	Leu 50		Val	Asp	Asn	Arg 55	Gly	His	Gly	Arg	Ser 60	ДЗÞ	Lys	Pro	Leu
Gly 65		Asp	Phe	Tyr	Arg 70	Phe	Glu	Aŝ'n	Phe	Ile 75	Ser	Аар	Leu	Aap	Ala 80
Val	Val	Arg	Glu	Thr 85	Gly	Val	Glu	Lys	Phe 90	Val	Leu	Val	Gly	His 95	Ser
Phe	Gly	Thr	Met 100	Ile	Ser	Met	Lys	Tyr 105	Сув	Ser	Glu	Tyr	Arg 110	Asn	Arg
Val	Leu	Ala 115	Leu	Ile	Leu	Ile	Gly 120	Gly	Gly	Ser	Arg	Ile 125	Lys	Leu	Leu
His	Arg 130	Ile	Gly	Tyr	Pro	Leu 135	Ala	Lys	Ile	Leu	Ala 140	Ser	Ile	Ala	Tyx
Lys 145	Lys	Ser	Ser	Arg	Leu 150	Val	Ala	Asp	Leu	Ser 155	Phe	Gly	Lys		Ala 160
Gly	Glu	Leu	Lys	Glu 165	Trp	Gly	Trp	Lys	Gln 170	Ala	Met	Asp		Thr 175	Pro

Ser Tyr Val Ala Met Tyr Thr Tyr Arg Thr Leu Thr Lys Val Asn Leu 190

Glu Asn Ile Leu Glu Lys Ile Asp Cys Pro Thr Leu Ile Ile Val Gly 195

Glu Glu Asp Ala Leu Leu Pro Val Ser Lys Ser Val Glu Leu Ser Arg 210

Arg Ile Glu Asn Ser Lys Leu Val Ile Ile Pro Asn Ser Gly His Cys 235

Val Met Leu Glu Ser Pro Ser Glu Val Asn Arg Ala Met Asp Glu Phe 245

Ile Ser Ser Ala Gln Phe

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 251 AMINO ACIDS
 - (B) TYPE: AMINO ACID

5

10

20

25

30

35

40

45

50

55

- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Leu His Ile Leu Leu Arg Ser Phe Phe Leu Ala Val Arg Ser Asn Ser 195 200 205

Glu Lys Arg Lys Glu Phe Cys Asp Leu Val Ile Val Pro Glu Leu Glu 210 215 220

Glu Phe Thr Pro Leu Asp Val Arg Lys Ala Asp Gln Ile Met Glu Arg 225 230 235 240

Gly Tyr Ile Lys Ala Leu Glu Val Leu Ser Glu 245 250

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 297 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN

45

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

10	Met 1	. Phe	Ast	Ile	Asn 5	Val	Phe	· Val	. Aso	Ile 10		Trp	Leu	Тух	Phe 15	Ser
	Gly	Ile	Val	Met 20		The	Val	Glu	Glu 25		Ala	Leu	Leu	G]u 30		Gly
15	Val	Arg	Val 35		Tyr	Arg	Сув	Val 40		Pro	Glu	Lys	Ala 45		Asn	Thr
	Leu	Ile 50	Ile	Gly	Ser	His	Gly 55	Leu	Gly	Ala	His	Ser 60	Gly	Ile	Tyr	Ile
20	Ser 65	Val	Ala	Glu	Glu	Phe 70		Arg	His	Gly	Phe 75	Gly	Phe	Сув	Met	His 80
	Ąsp	Gln	Arg	Gly	His 85	Gly	Arg	Thr	Ala	Ser 90	Asp	Arg	Glu	Arg	Gly 95	Tyr
25	Val	Glu	Gly	Phe 100	His	Asn	Phe	Ile	Glu 105	qaA	Met	Lys	Ala	Phe 110	Ser	Авр
	Tyr	Ala	Lys 115	Trp	Arg	Val	Gly	Gly 120	qaA	Glu	Ile	Ile	Leu 125	Leu	Gly	His
30	Ser	Met 130	Gly	Gly	Leu	Ile	Ala 135	Leu	Leu	Thr	Val	Ala 140	Thr	Tyr	Lys	Glu
	Ile 145	Ala 	Lys	Gly	Val	Ile 150	Ala	Leu	Ala	Pro	Ala 155	Leu	Gln	Ile	Pro	Leu 160
35	Thr	Pro	Ala	Arg	Arg 165	Leu	Val	Leu	Ser	Leu 170	Ala	Ser	Arg	Leu	Ala 175	Pro
	His	Ser	Lys	Ile 180	Thr	Leu	Gln	Arg	Arg 185	Leu	Pro	Gln	Lys	P ro 190		Gly
40	Phe	Gln	Arg 195	Ala	Lys	qeA	Ile	Glu 200	Tyr	Ser	Leu		Glu 205	Ile	Ser	Val
		Leu 210	Val	Asp	Glu :	Met	Ile 215	Lys	Ala	Ser :		Met 220	Phe	Trp	Thr	Ile

Ala Gly Glu Ile Asn Thr Pro Val Leu Leu Ile His Gly Glu Lys Asp 240

Asn Val Ile Pro Pro Glu Ala Ser Lys Lys Ala Tyr Gln Leu Ile Pro Ser Phe Pro Lys Glu Leu Lys Ile Tyr Pro Asp Leu Gly HisoAsn Leu Phe Phe Glu Pro Gly Ala Val Lys Ile Val Thr Asp Ile Val Glu Trp Val Lys Asn Leu Pro Arg Glu Asn Pro 295

191	INIEODI	4 ATION!		SEV.	ın	NIU・38
141	INFORM		CIT	SEG	JU	140.00

(i) :	SEQl	JENCE	CHAR	ACTE	risti	CS
-------	------	-------	------	------	-------	----

- (A) LENGTH: 262 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN

10

20

25

35

40

45

50

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

> Glu Gly Ser Arg Arg Leu Phe Glu Glu Leu Ala Val Glu Asn Lys Thr 210 215 220

> Leu Arg Glu Phe Glu Gly Ala Tyr His Glu Ile Phe Glu Asp Pro Glu 225 230 235 240

Trp Ala Glu Glu Phe His Glu Thr Ile Val Lys Trp Leu Val Glu Lys 245 250 255

Ser Tyr Ser Ser Ala Gln 260

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 249 AMINO ACIDS
 - (B) TYPE: AMINO ACID

5

10

45

50

55

- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Leu Ile Gly Asn Leu Lys Leu Lys Arg Phe Glu Glu Val Asn Leu Val 1 5 15 Leu Ser Gly Gly Ala Ala Lys Gly Ile Ala His Ile Gly Val Leu Lys 20 25 30 15 Ala Leu Glu Glu Leu Gly Ile Lys Val Lys Arg Leu Ser Gly Val Ser Ala Gly Ala Ile Val Ser Val Phe Tyr Ala Ser Gly Tyr Thr Pro Asp
50
60 20 Glu Met Leu Lys Leu Lys Glu Val Asn Trp Leu Lys Leu Phe Lys
-65 70 75 80 Phe Lys Thr Pro Lys Met Gly Leu Met Gly Trp Glu Lys Ala Ala Glu 85 90 95 25 Phe Leu Glu Lys Glu Leu Gly Val Lys Arg Leu Glu Asp Leu Asn Ile 100 105 110 Pro Thr Tyr Leu Cys Ser Ala Asp Leu Tyr Thr Gly Lys Ala Leu Tyr 115 120 30 Phe Gly Arg Gly Asp Leu Ile Pro Val Leu Leu Gly Ser Lys Ser Ile 130 140 Pro Gly Ile Phe Glu Pro Val Glu Tyr Glu Asn Phe Leu Leu Val Asp Gly Gly Ile Val Asn Asn Leu Pro Val Glu Pro Leu Glu Lys Phe Lys Glu Pro Ile Ile Gly Val Asp Val Leu Pro Ile Thr Gln Glu Arg Lys 40 Ile Lye Asn Ile Leu His Ile Leu Ile Arg Ser Phe Phe Leu Ala Val 195 200 205 Arg Ser Asn Ser Glu Lys Arg Lys Glu Phe Cys Asn Val Val Ile Glu 210 220

> Pro Pro Leu Glu Glu Phe Ser Pro Leu Asp Val Asn Lys Ala Asp Glu 225 230 235 236

Ile Phe Cys Gly Asp Met Arg Ala Leu 245

- (2) INFORMATION FOR SEQ ID NO:40:
 - (i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 339 AMINO ACIDS (B) TYPE: AMINO ACID
- (B) TYPE: AMINO ACID
 (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN

5

50

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

10		Met 1	Pro	Ala) Asn	qaA 5	Sex	Pro	Thi	: Ile	Asp 10		Ast	Pro	Arg	Gly	/ Ile
		Lev	Arg	Asi	Ala 20		Ala	Gln	Va]	11e 25		Ala	Thr	Ser	Gly 30	Leu	Arg
15		Lys	Ala	Phe 35	Leu	Lys	Arg	Thr	His 40		Ser	Tyr	Leu	Ser 45		Ala	Gln
		Trp	Leu 50		Leu	Asp	Ala	Gly 55	Asn	Gly	Val	Thr	Leu 60		Gly	Glu	Leu
20		Asn 65		Ala	Pro	Ala	Thr 70		Ser	Ser	Ser	His 75	Pro	Ala	His	Lys	Asn 80
		Thr	Leu	Val	Ile	Val 85	Leu	His	Gly	Trp	Glu 90	Gly	Ser	Ser	Gln	Ser 95	Ala
25		Tyr	Ala	Thr	Ser 100	Ala	ĠĮĄ	Ser	Thr	Leu 105		Asp	Asn	Gly	Phe 110	_	Thr
		Phe	Arg	Leu 115	Asn	Phe	Arg	Asp	His 120	Gly	Asp	Thr	Tyr	His 125	Leu	Asn	Arg
30		Gly	Ile 130	Phe	Asn	Ser	Ser	Leu 135	Ile	Asp	Glu	Val	Val 140	Gly	Ala	Val	Lys
	Ç	Ala 145	Ile	Gln	Gln	Gln	Thr 150	Asp	Tyr	Asp	Lys	Tyr 155	CAa	Leu	Met	Gly	Рће 160
35		Ser	Leu	Gly	Gly	Asn 165	Phe	Ala	Leu	Arg	Val 170	Ala	Val	Arg	Glu 175	Gln	His
		Leu	Ala	Lys 180	Pro	Leu	Ala	Gly	Val 185	Leu	Ala	Val	Cys	PTO 190	Val	Leu	Asp
40		Pro	Ala 195	His	Thr	Met	Met	Ala 200	Leu	Asn	Arg		Ala 205	Phe	Phe	Tyr	Gly
		Arg 210	Tyr	Pbe	Ala		L ув 215	Trp	Lys	Arg	Ser	Leu 220	Thr	Ala	Lys	Leu	Ala 225
45		Ala	Phe	Pro		Tyr 230	Lys	Tyr	Gly	Lys	Авр 235	Leu	Lys	Ser		H18 240	Thr
		Leu	Asp		Leu . 245	Asn .	Asn	Tyr		Ile 250		Arg	Tyr	Thr	Gly :	Phe	Asn

							EP 0	880 5	90 B	1						
	Sex	Val	Ser 260	Glu	Tyr	Phe	Lys	Ser 265	Tyr	Thr	Leu	Thr	Gly 270	Gln	Lys	Leu
5	Ala	Phe 275	Leu	asa	Сув	Pro	Ser 280	Tyr	Ile	Leu	Ala	Ala 285	Gly	Asp	Asp	Pro
Á	Ile 290	Ile	Pro	Ala	Ser	Авр 295	Phe	Gln	Lys	Ile	Ala 300	Lys	Pro	Ala	Asn	Leu 305
10	His Ile Thr Val Thr Gln Gln Gly Ser His Cys Ala Tyr Leu Glu 310 315 320 Leu His Lys Pro Ser Ala Ala Asp Lys Tyr Ala Val Lys Leu Phe 325 330 335															Asn
	Leu His Lys Pro Ser Ala Ala Asp Lys Tyr Ala Val Lys Leu Phe 325 330 335															Gly
15																
	(2) INFORMA	TION I	OR S	EQ ID	NO:4	11:										
20	(i) SEQUI	ENCE	CHAF	RACTE	RIST	ics										
25	(B) T	ENGTI YPE: A OPOLO	MINC	ACID)	IDS					•					
25	(ii) MOLE	CULE	TYPE	: PRO	TEIN											
	(xi) SEQL	JENCE	DES	CRIPT	ION:	SEQI	D NO:	41:								

	Met 1		Asp	Met	Pro 5	Ile	Asp	Pro	Val	Tyr 10		Gln	Leu	Ala	Glu 15	Tyr
	Phe	. Asp	Ser	Leu 20		Lya	Phe	Ąsp	Gln 25	Phe	Ser	Ser	Ala	Arg 30	Glu	Tyr
	Arg	Glu	Ala 35	Ile	Asn	Arg	Ile	Туг 40	Glu	Glu	Arg	Asn	Arg 45	Gla	Leu	Ser
)	Gln	His 50	Glu	Arg	Val	Glu	Arg 55	Val	Glu	Asp	Arg	Thr 60	Ile	Lya	Gly	Arg
•	Asn 65		đeA	Ile	Arg	Val 70	Arg	Val	Tyr	Gln	Gln 75	ГЛа	Pro	Asp	Ser	Pro 80
.	Val	Leu	Val	Tyr	Tyr 85	His	Gly	Gly	Gly	Phe 90	Val	Ile		Ser		Glu
,	Ser	His	yab	Ala 100	Leu	Cys	Arg	Arg	Ile 105	Ala	Arg	Leu	Ser	Asn 110	Ser	Thr
	Val	Val	Ser 115	Val	Авр	Týr	Arg	Leu 120	Ala	Pro	Glu	Нiá	Lys 125	Phe	Pro	Ala
,	Ala	Val 130	тут	Ąsp	Cys	Tyr	Asp 135	Ala	Thr	Lys	Trp	Val 140	Ala	Glu	Asn	Ala
	Glu 145	Glu	Leu	Arg	Ile	Asp 150	Pro	Ser	Lys	Ile	Phe 155	Val	Gly	Gly		Ser 160
5	Ala	Gly	Gly		Leu 165	Ala	Ala	Ala		Ser 170	Ile	Met	Ala		Asp 175	Ser
	Gly	Glu	Asp	Phe 180	Ile	Lys	His	Gln	11e 185	Leu	Ile	Tyr		Val 190	Val .	Asn
0	Phe	Val	Ala	Pro	Thr	Pro	Ser .	Leu	Leu (Glu	Phe	Gly (Glu (Gly :	Leu 1	Trp
		•	195					200)				205	:		
5	Ile	Leu 210		Gln	Lys	Ile	Met 215	Sez		Phe	Ser	Glu 230	Gln		Phe	Ser
	Arg 235	Glu	G1u	Asp	Lys	Phe 240	Asn	Pro	Leu	Ala	Ser 245	Val		Phe	Ala	Asp 250
0	Leu	Glu	Aøn	Leu	Pro 255	Pro	Ala	Leu	Ile	Ile 260	Thr	Ala	Glu	Тух	Asp 265	
	Leu	Arg	Asp	Glu 270	Gly	Glu	Val	Phe	Gly 275		Met	Leu	Arg	Arg 280		Gly
5	Val	Glu	Ala 285	Ser	Ile	Val	Arg	Tyr 290	Arg	Gly	Val	Leu	His 295		Phe	Ile
	Asn	Tyr 300	Tyr	Pro	Val	Leu	Lys 305	Ala	Ala	Arg	Двр	Ala 310		asa	Gln	Ile
0	Ala 315		Leu	Leu	Val	Phe 320										

(2) INFORMATION FOR SEQ ID NO:42:

55

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 305 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN

35

40

45

55

- (xi) SEQUENCE DESCRIPTION: SEO ID NO:42:
- 10 Met Pro Leu Asp Pro Arg Ile Lys Lys Leu Leu Glu Ser Ala Leu Thr
 15 Ile Pro Ile Gly Lys Ala Pro Val Glu Glu Val Arg Lys Ile Phe Arg
 20 Pro Lys Ile Pro Gly Ser Glu Thr Val Ile Asn Ala Arg Val Glu Asp
 65 Pro Lys Ser Ser Gly Pro Tyr Gly Val Leu Val Tyr Leu His Gly Gly
 65 Phe Val Ile Gly Asp Val Glu Ser Tyr Asp Pro Leu Cys Arg Ala
 85 Pro Glu Tyr Lys Phe Pro Ser Ala Val Ile Asp Ser Phe Asp Ala
 116 Pro Glu Tyr Lys Phe Pro Ser Ala Val Ile Asp Gly Lys Met Gly
 127 Val Ala Ile Ala Gly Asp Ser Ale Gly Gly Asn Leu Ala Ala Val Val
 145 Val Ala Ile Ala Gly Asp Ser Ale Gly Gly Asn Leu Ala Ala Val Val
 145 Pro Gly Asp Ser Ale Gly Gly Asn Leu Ala Ala Val Val
 146 Pro Gly Asp Ser Ale Gly Gly Asn Leu Ala Ala Val Val
 146 Pro Gly Asp Ser Ale Gly Gly Asn Leu Ala Ala Val Val
 146 Pro Gly Asp Ser Ale Gly Gly Asn Leu Ala Ala Val Val
 146 Pro Gly Asp Ser Ale Gly Gly Asn Leu Ala Ala Val Val

	Ala	Leu	Leu	Ser	Lys 165	Gly	Lys	Ile	Asn	Leu 170	Lys	Tyr	Gln	Ile	Leu 175	Val
5	Tyr	Pro	Ala	Val 180	Ser	Leu	Asp	Asn	Val 185	Ser	Arg	Ser	Met	Ile 190		Tyr
	Ser	Ąsp	Gly 195	Phe	Phe	Leu	Thr	Arg 200	Glu	His	lle	Glu	Trp 205	Phe	Gly	Ser
10	Gln	Tyr 210	Leu	Arg	Ser	Pro	Ala 215	qeA	Leu	Leu	Asp	Phe 220	Arg	Phe	Ser	Pro
	Ile 225	Leu	Ala	Gln	Asp	Phe 230	aea	Gly	Leu	Pro	Pro 235	Ala	Leu	Ile	Ile	Thr 240
15	Ala	Glu	Tyr	Авр	Pro 245	Leu	Arg	Ąsp	Gln	Gly 250	Glu	Ala	Tyr	Ala	Asn 255	Lys
	Leu	Leu	Gln	Ala 260	Gly	Val	Ser	Val	Thr 265	Ser	Val	Arg	Phe	Asn 270	Asn	Val
20	Ile		Gly 275	Phe	Leu	Ser	Phe	Phe 280	Pro	Leu	Met	Glu	Gln 285	Gly	Arg	Asp
		Ile 290	Gly	Leu	Ile		Ser 295	Val	Leu	Arg		Val 300	Phe	Tyr	Asp	Lys
25	Ile 305															

(2) INFORMATION FOR SEQ ID NO:43:

30

35

45

50

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 605 NUCLEOTIDES
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: GENOMIC DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

					T AJ								Gly				t S			48
5					ng co											r G)				96
• .				u AE	C TI				sn T							g As				144
10				r (3)	IA AI			lu Pi							Se					192
	•		CA	C TC	T AT		y Gi	CT	ra g					Tyz				A A		240
15					A AG o Se		B H1										y Th			288
					C AG y Se: 10	r Hi				lu L							y Ph			336
20																				
		1	CTA Leu	TTA Leu	Leu	Lys	AAT Aen	AGC	GT Val	CI	u Pl	TT 1	MTA Leu	CTC Leu	TCT Ser	Lys	ABI	GO	CGAT	384
25		7	LTD LGQ	CCT Pro 130	TII Phe	AAA	GCC	AAG Lys	CTA Leu 135	יענ	r ac	C A	TT :	Ala	GGC Gly 140	125 GAC Asp	TTA	CCC	ATT	432
3 <i>0</i>		G	igc ily .45	TTA	ATG Mat	CCA Pro	CTC Leu	ATT 110 150	GTA Val	. AAJ	A GG	K A y S	er i	ogc ·	TCT	gat Asp	GC GC	ACT The	GTA Val 160	480
		ī	TO	CTA Leu	cat Asp	GAA Glu	ACC Thr 165	aag Lyb	CTA Leu	AAC Lys	Gl	у М	TG (et ; 70	CT (Blu Blu	CAC His	aag Lyb	GTG Val 175	TTT	528
35		Н	AT i	TTA Leu	AGC Ser	CAT His 180	ACA Thr	AGT Ser	ATG Met	ATI Ile	TA: Ty: 18	T S	CT C	oc (ZAA Blu	Val	GTT Val 190	AAT Asn	TAT Tyr	576
				Leu	GAG Glu 195						11		•							605
40																				
	(2) INFORM	MATIC)N F	OR	SEQ	ID N	O:44	:												
	(i) SEC	UEN	CE (CHA	RAC	TERI	STIC	s												
45	(B) (C)	LENO TYPI STR TOP	E: N	UCL DEDI	EIC NESS	ACID S: SII)													
	\				-	-														

50

(ii) MOLECULE TYPE: GENOMIC DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

50

	ATG Met 1	ATA Ile	lyb	AAC Asn	Pha 5	GAC Asp	Arg Arg	GAA Glu	TAA Asn	TCT Ser 10	Ser	TTA	Val	Leu	Ser 15	Gly	48
5	GOT	GGT Gly	GCT Ala	CTG Leu 20	GGT Gly	ATT 1le	GCT Ala	CAC His	TTG Leu 25	Gly	GTA Val	CTG	CAT His	GAC Asp 30	CTT	GAA Glu	96
	lys	CAA Gln	AAT Asn 35	ATT Ile	OTA Val	CCA Pro	AAT Asn	GAA Glu 40	ATT Ile	GTT Val	Gly GGT	ACA Thr	AGT Ser 45	ATG Met	Gly	Gly Gly	144
10	Ile Ile	Ile	CJA	GCA Ala	ser	ATG Met	GCT Ala 55	ATC Ile	GGG	ATG Met	AAA Lys	eo eya eya	AAA Lys	GAA Glu	ATA Ile	CTC	192
	GAA Glu 65	G) n Gyr	ATC Ile	AAA Lys	AAC ABB	711 Phe 70	TCC Ser	aat Asd	GTC Val	TTC. Phe	AAC ABD 75	TGG TEP	ATA Ile	aaa Lys	TTC Phe	TCT Ser 80	240
15	TTT Phe	Ser	Gly	Aen	Ser 85	Val	Va)	Asp	Asn	G1u 90	Lys	Ile	Ala	Lys	Ile 95	Phe	288
÷	GAT Asp	ACT Thr	Leu	TTT Phe 100	AAA Lys	GAC Asp	AGA Arg	Lys	ATG Met 105	ACA Thr	GAT Asp	ACG Thr	Val	ATC Ile 110	CCT Pro	CII Leu	336
20	AAA Lys	Leu										Lya					384
	Ala :	rcg Ser 130	gat Asp	GAT Asp	GTA Val	Leu	ATC Ile 135	AAA Lys	GAT Asp	GCA . Ala	Ile :	CTC Leu 140	TCA Ser	ACA Thr	ATG Met	GCA Ala	432
	ATA (Ile) 145	Pro	GGT Gly	GTA '	Phe (GAA (Glu (150	GAA Glu	CAT His	ATT . Ile	Ile :	GAT Asp 155	GCT (GAA Glu	ACC Thr	Tyr	GGC Gly 160	480
30	ABI	o Gl	y Ph	e Le	16:	9 G1: S	u As	n Le	u Gl	y V a 17	l As: 0	n Gl	u Al	a Th	r Ph 17	-	528
35	GA1 Ass	r GT Va	TTI Lei	A GC: 1 Al: 180	va.	A GA: L As ₁	r Gr	C ATN 1 Mei	3 GG 5 G1: 18:	i GTI	AA E IBA L	C TC	T TT	7 GA e Gl 19	u Ly	A GCA 8 Ala	576
	ATC Met	Pro	GAG Asp 199) ASI	TTO Phe	Phe	Ly:	A ACI Thi 200	: Sei	A AA?	r GT	Le	A GAI 3 G1: 20:	a Ne	G II	r GAA e Glu	624
40	AAA Lys	Ser 210	. Met	Arg	Leu	Phe	11e 215	: Iyı	AAC TRA	CAC Glr	ACI Thr	Glr 220	Thi	CA:	I AT	Lys	672
	AAT Asn 225	Ala	TAA J nea j	Lys Lys	TAA GBA	ATT Ile 230	Tyr	CTI Leu	ATT Ile	GJ: GY)	235	Val	ACC	Lys	GAC Glu	TAT Tyr 240	720
45	aaa Lyb	ACA Thr	TTT Phe	CAA Gln	777 Phe 245	CAT Ris	Lyn	CAT His	aaa Lys	GA9 Glu 250	Ile	Arg	GCI	Leu	GGC Gly 255	TTG Leu	768
			CTG Leu	TG													779

(2) INFORMATION FOR SEQ ID NO:45:

50

55

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 905 NUCLEOTIDES
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: GENOMIC DNA

·

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

								Leu			PXO	48
10		gac asp								Asp	TTA Leu	96
		TTC										244
15		CCC Pro 50										192 -
		AAG Lys										240
20		TTC Phe										288
		GAA Glu										336
25		CAC His										384
	Trp	ATC Ile 130			Gly							432
30 .	GCC Ala 145			Asp .			Aen			Val		480

301

					Gln				Val				CTA	528
5				Val						Pro			GAG Glu	576
										ATG Met			GGG GGG	624 -
10										TCC Ser 220				672
15										TCA Ser				720
										ACC Thr				768
20	CTA Leu		G1u				Ser			AGA Arg	Tyr			826
	ATT Ile	His				Phe				Thr				864
	GCC Ala				Ala -				Ser			Ser	Val	905
	Leu													

(2) INFORMATION FOR SEQ ID NO:46:

30

35

45

50

55

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 978 NUCLEOTIDES
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: GENOMIC DNA
- 40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

.× ·.

		GCC Ala															48
		AGG Arg															96
		GAA Glu															144
		CAT His 50															192
	AGG Arg 65	GGG Gly	CAG Gln	AGA Arg	ATA Ile	AAG Lys 70	GGC Gly	TGG Trp	CTT Leu	CTT Leu	GTT Val 75	CCG PTO	aag Lyb	TTG Leu	GCG Ala	GAA Glu 80	240
•		aag Lyb															288
	G17 GGI	TII Phe	Pro	CAC His 100	gac Asp	TGG TTP	CTG Leu	TTC Phe	TGG Trp 105	CCG Pro	TCA Ser	ATG Met	GOT Gly	TAC Tyr 110	ATC Ile	TGT Cys	336

	Phe	CTC Val	ATG Met	. Ant	: ACC	AG(3 GG	3 CA 7 G1: 12	n Gly	A AGO y Set	r Gly	TI	ATO Met 135	Lye	GI)	A GAC	384
5	ACA	130	qaA (TAC	Pro	Glu	3 667 1 61 ₇ 135	PEC	A GT(GA1	Pro	CAG G1: 140	זעז י	Pro CCC	GI)	TTC Phe	432
	Met 145	Thr	AGG Arg	GGC	Ile	CTG Lev 150	yel	Pro	GC3 OC1)	ACC Thi	TAT TYT 155	Tyz	Tyr	AGG Arg	CG2 Arg	GTC Val 160	480
10	TTC Phe	GTG Val	GAT Asp	GCG Ala	Val 165	Arg	GCG Ala	GTG Val	GAN Glu	GCA Ala 170	GCC Ala	ATT	TCC	TTC	Pro 175	Arg	528
	GTG Val	GAT Asp	TCC Ser	AGG Arg 180	AAG Lys	GTG Val	GTG Val	GTG Val	GCC Ala 185	Gly	egc egc	agt Ser	CAG Oln	01y 037	gra GCG	GGA	576
15	ATC Ile	CCC	CII Leu 195	GCG Ala	GTG Val	AGT Ser	Ala	CTG Leu 200	TCG Ser	AAC Asn	AGG Arg	GTG Val	AAG Lys 205	GCT Ala	CTG Leu	CTC Leu	624
. 20											AGG Arg						672
20											TTC Phe 235						720
25	OAD qe4			Glu							TCC Ser			App			768
			Ala								CTG Leu		Ser				816
30	ATG Met.	qaA					Pro					Ala.					864
	TYT					Glu					Pro '						912
35	305 G1y	Gly (3ly :	Ser 1	Phe	Gln 310				Gln					Lys :		960
	CTA 1			31u (TAG						٠.					978

(2) INFORMATION FOR SEQ ID NO:47:

40

45

50

55

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 879 NUCLEOTIDES
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: GENOMIC DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

ATG CGC ACC CTC TCC TTC GGT CCG ATG ACC ACA GGG GGA AGC ATT CAC Met Arg Thr Leu Ser Phe Gly Pro Met Thr Thr Gly Gly Ser Ile His 1

· 5	ATG G Net A		hr M					rg G					n Ax				96
10 .			y Al					a As					la Ly			ro occ	
15	AF	g Ma 5	t 01 0	y Ty	t yi	a Al	a Ly	19 GJ 5	y al	.a Va	il Ty	T A)	a I)	e Il	e CI	C AGC	192 240
	Let 65	a Ala	a Lei	u Ly	B Lei	a Al	a Al	a Gl	y 61	n G1	y G1:	y Ar 5	g Th	r Th	x As	P Ser 80	
20	His	G1)	y Ala	val	L Ası 95	7 Thu	r Va.	l al	s Hi	9 61	y Pro) Ph	. G1;	y Va	1 AL 9:	a Leu 5	288
	Leu	Ala	Val	Lev 100) 1 Val	Va]	l Gly	/ Lei	10:	2 G1;	y Tys	· Val	l Vai	11))	Phe	336
25	Ala	G1n	115	Phe	Val	. Ast	The	120) 1 Ası	Lys	a) y	! Sex	125	Ala	Ly:	GOA Gly	384
	Ile	130	Thr	Arg	Ala	Met	135	Phe	: Leu	Sex	• 01 y	140	lle	نارو	Ala	Ser Ser	432
30	145	Ala	Phe	Phe	Ala	Ala 150	Gln	Ser	Leu	Val	Gly 155	Ala	Ala	His	Gly	160	480
35	Ser	Lys	GGG	Thr	Gln 165	Gly	ırp	Thr	Ala	Thx 170	Leu	Me:	Glu	Glh	Pro 175	Phe	528
	Gly	Arg	Val	Leu 180	Val	Ala	Leu	Val	01y 185	Leu	GIA	116	Val	730 GJA	Phe	Ala	576
40	Leu	Lys	CAG Gln 195	Phe	Hib	Thr	Ala _,	11p 200	Lys	Ala	Lys	Phe	Arg 205	Glu	Lys	Leu	624
	Thr	Leu 210	ACC	Gly	Leu	Ala	Ala 215	Arg	ГÀВ	Gln	Ris	His 220	Iļe	Glu	Arg	Met	672
45	Cys (225	Gln		Gly	Ile	Ala 230	Ala	Arg	Gly	Val	Val 235	Phe	Ala	Val	Ile	Gly 240	720
	Gly i	Pho .	Leu '	Val .	Arg . 245^	Ser	Ala	Val	Авр	Ala 250	Asn	Pro	01y	ejn.	Ala 255 '	Lys	768
50	GIV I	Leu (01y	31u 2 260	Ala 1	Leu .	Ala	Val	Val 265	Ala .	Arg (Gla	Pro	9er 270	Gly .	Asp	916
	Val I	Leu I	leu (275	ely (Val V	Val i	Ala :	Ala 280	Gly	leu '	Val 1	la '	Tyr 1 285	Ala :	Ala	TYE TYE	864
55	CTG 1					urg :					LAG.						897

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 914 NUCLEOTIDES

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: GENOMIC DNA

•	(xi) SEQUENCE DESCRIPTION: SEQ. ID NO:48:	
5 ·	Met Ser Lys Phe Ala He Leu Trp Ala Leu He Thr Ala Tyr Leu Pro 1 15	8
	GRA CCT GTG ATG AAA CTG GTA TAT TTA GGG CGC GGA ACG CTT GGG 900 Olu Pro Val Met Lys Leu Val Tyr Leu Gly Arg Arg Glu Thr Leu Gly 20 25 30	6
10	GCA CGG ACG CTT GAC GTT AAA GCC CAA GCT GTC GGG CGG CTG GCC AAT 140 Ala Arg Thr Leu Asp Val Lys Ala Gln Ala Val Gly Arg Leu Ala Asn 35 40 45	3
	GCA ACA AGA CCT GTC GGG GTG ATT CCG ACG GTC GAG GAA AGC CGG AAG Ala Thr Arg Pro Val Gly Val Ile Pro Thr Val Glu Glu Ser Arg Lyb 50 60	!
15	ATG ACG GAT AAA GCC GTT AGC CTT TTT GAT CAG CCC GCC CCC GAA TTA 240 Met Thr Asp Lys Ala Val Ser Leu Phe Asp Gin Pro Ala Pro Glu Leu 65 70 75 80	1
	TTC CGT AAA AAA GAC ATT CAG ATT GAC GGG GCT GAA GGG CCT ATT GAT 288 Phe Arg Lys Lys Asp Ile Gln Ile Asp Gly Ala Glu Gly Pro Ile Asp 95 90 95	
20	GCC CGT ATT TAC AGC GGC CCT GCA ANA CAT CGC CCA CGA CCA ATA CTA Ala Arg Ile Tyr Ser Gly Pro Ala Lys His Arg Pro Arg Pro Ile Leu 100 105 110	
	GTG TAT TTT CAC GGC GGT GGC TGG GTT CAG GGC AAT CTG GAC AGC CAT Val Tyr Phe His Gly Gly Gly Trp Val Gln Gly Asn Leu Asp Sar Ris	
25	GAC GGG GIT TGC GGC AAG CTU GCA AAA TGG GCG AAC TGC ATT GIT ATC ASP Gly Val Cys Gly Lys Leu Ala Lys Trp Ala Asn Cys Ile Val Ile 130 140	
30	TCG GTC GAT TAT CGT CTA GCG CCC GAA CAC AAA TTT CCT TGT GCG CCG Ser Val Asp Tyr Arg Leu Ala Pro Glu His Lys Phe Pro Cys Ala Pro 145 150 155 160	
	CTT GAT GCG ATT GCG GCC TAT AAA TGG GTG CCC GCC AAC GCA ACA AAC 528 Leu Asp Ala Ile Ala Ala Tyr Lys Trp Val Arg Ala Asn Ala Thr Asn 165 170 175	
35	CTT GGC GGC GAT CCT GAA CGT ATC GGC GTT GGC GGC GAT AGC GCA GGG 576 Leu Gly Gly Asp Pro Glu Arg Ile Gly Val Gly Gly Asp Ser Ala Gly 185 190	
	GGC AAT CTT GCC GCC GTT GTC TGC CAA CAA ACC GCC ATG AAC GGC GAO 624 Gly Asn Leu Ala Ala Val Val Cys Gln Gln Thr Ala Met Asn Gly Glu 195 200 205	
40	CGC ACA CCA GAT CTG CAA GTC CTG ATC TAT CCG GCG CTG GAT GCA CGC 672 Arg Thr Pro Asp Leu Gln Val Leu Ile Tyr Pro Als Leu Asp Als Arg 210 220	
	ATG ATC TCG ACC TCG ATG GAG GAA TTG CGT GAT GCC TAC ATC TTG CCG Net Ile Ser Thr Ser Met Glu Glu Leu Arg Asp Ala Tyr Ile Leu Pro 230 235 240	
45	AAA TCC AGA ATG GAG TAT TTC CTC GGC CTA TAT ACG CGT GGC CCT GAC LyB Ser-Arg Met Glu Tyr Phe Leu Gly Leu Tyr Thr Arg Gly Pro Asp 245 250 255	
	GAT ATC GAG GAC CTT AGG ATG TCG CCA ATT CTC AGG GAT ACC GTC GCG ASP Ile Glu Asp Leu Arg Met Ser Pro Ile Leu Arg Asp Thr Val Ala 260 265 270	
50	SAT CAA CCC CAA GCC TGC ATT GTC ACC TGT GGG TTT GAC CCT GCG CGA 864 Asp Gln Pro Gln Ala Cys Ile Val Thr Cys Gly Phe Asp Pro Ala Arg 275 280 285	
	CCA CGG GAA CAC CTA CGC CGA ACG CTT AAT TGC CGA GGG GAT AGA CGT 912 Arg Arg Glu His Leu Arg Arg Thr Leu Ash Cys Arg Gly Asp Arg Arg	

(2) INFORMATION FOR SEQ ID NO:49:

55

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 926 NDCLEOTIDES
- (B) TYPE: NUCLSIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

10

15

20

25

30

35

40

45

50

- (ii) MOLECULE TYPE: GENOMIC DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

	Va.	3 AGK L Sei L	C AT	r CO	Cro Lev	YL	Lev	TT)	AAC ABE	TIG) Phe	Lev	AA1	r ACt	Phi	T GAA 8 Glu 5	48
	Lys	CCI Pro	Lys	Let 20	a Ala	GOG Ala	GCC Ala	Lye	ACG Thr 25	Pro	yet Gy3	GAI Asp	Leu	CGI Arg 30	, Lýs	A TCG Ser	96
	Phe	GAP Glu	Leu 39	ı Lye	GCG Ala	AGG Arg	Phe	Leu 40	Pho	Pro	GCG	CCA Pro	CGI Arg	Lys	ACI Thi	A AGG C Arg	144
	Phe	AGT Set	Hie	r gat B Abp	GTA Val	Leu	CAG Gln 55	3er	GCC	ATC	GGG	TCG Ser 60	GTA Val	AAT nea	GCC	Gln	192
	TGG TIT 65	ala (Lye	TCC Ser	Lys	TCT Ser 70	. yjs	TCT Ser	gat Asp	gac Asp	AGG Arg 75	gta Val	ATC	Leu	TAT	Phe 80	240
	CA1 His	GGG Gly	GCZ	GGG Gly	TAT Tyr 85	Val	Phe	Gly	TCA Ser	CCA Pro 90	aaa Lys	ACG Thr	CAC His	Arg	GCA Ala 95		288
	Leu	GCG	Arg	Leu 100	TCG Ser	GCA Ala	ATG Met	ACA Thr	GOT Gly 105	CTT	TCT Ser	GCG Ala	TGC Cys	CTT Leu 110	CCA Pro	GAT Asp	336
	TAT Tyx	AGG Arg	Leu 115	Ala	CCA Pro	GJ <i>n</i> GYG	CAC His	CCA Pro 120	TTT Phe	CCA Pro	GCC Ala	GCG Ala	ATC fle 125	GAA Glu	gat Asp	GCA Ala	384
					aaa Lyb												432
	ATT Ile 145	ATA Ile	CTG Leu	GGG Gly	gjà egg	GAC A8p 150	AGT Ser	GCT Als	gjy Get	OJA GCC	GGT Gly 155	TTG Leu	GTT Val	CTT Leu	GCT Alm	TTG Leu 160	480
					AAG Lys 165				Leu								528
					TTG Leu								Leu				576
	aaa Lys						Val					Ser .					624
	ATG Met					Leu .					Ala .						672
	TCG Ser 225				Ala					Met					Leu		720
:	GCA Ala	AGT Ser	GAC Asp	Ser	GAA Glu 245	ATC (CTG :	TTG (Leu .	Asp i	SAT : Asp (250	rge (Cys)	Leu I	ogg j larg l	Met 1	SCG (Ala : 255	gat Asp	768
	CAC His	TTG Leu	Arg	GCG Ala 260	CAA (Gln (GOT (GTC (Val '	Val '	01G # Val 1 265	CA (reb n	7.03 1	ile v	777 (Val (270	AA I	rac Red	816

					-														
•		CAT His	CCA Pro	CAT His 275	GTT Val	TGG Txp	CAT His	ATT	777 Phe 280	CAA Gln	CGC	CTT	CTA Leu	CCC Pro 285	GAA Glu	GCA Ala	CAT Asp		864
5		GJn	GGG Gly 290	CTG	CGG Arg	GCG Ala	ATT Ile	GCC Ala 295	GCG Ala	TGG Trp	ATT Ile	AAA Lys	CCT Pro 300	CTT Leu	TTA Leu	TCA Ser	GCT Gly	•	912
			AAC Asn			TA							•						926
10																			
	(2) INFORMAT	TION F	FOR	SEQ	ID N	O:50	:	,											
•	(i) SEQUE	ENCE	CHA	RAC [*]	TERI	STIC	S												
15	(A) LE	ENGT	H: 71	3 NU	CLE	DITC	ES												
	(B) T	YPE: N	NUCL	EIC	ACID														
	· (c) s																		
	(D) T	OPOL	OGY:	LINE	EAR														
20	(ii) MOLE	CULE	TYP	E: GE	ENON	AIC E	NA												
	(xi) SEQU	ENCE	DES	SCRI	PTIO	N: SI	EQ II	ON C	:50:										
25																			
30																			

	ATG Net 1	Let	Thi	A TII	AAT Asn 5	Val	TTA Leu	Tyz	G1)	ATC Met	Me	B AA	A CA	A AR	A CT S Le 1	y Al	ZA la	4	8
5	GCA Ala	AT7	Lev	ATG Met 20	Phe	TTA	GGG Gly	Leu	TCA Ser 25	. YJ9	A)	A GAG	GC AL	T CA a Gl	n As	C TG P Tz	G P	94	6
•	CCT Pro	GAC Asp	Leu 35	CAG Gln	AAA Lys	TAT Tyr	Arg	AGT Ser 40	GCT Ala	AAT ABD	Lys	Glu	4:	a Ly	A TTZ s Let	A CT	T U	144	•
1	Pro	Lys 50	Glu	AAC ABD	Arg	Lys	Val S5	Val	Phe	Met	01)	ABN 60	Se:	r Il	e Thi	. GJ	n	192	2
3	3CC 11a 65	TGG Trp	ATT	AGT Ser	CAG Gln	CGA Arg 70	Pro	GAG Glu	Phe	Phe	AG1 Sez 75	: Glu	AA: Ası	r GG n Gl	y Phe	TA ?	e	240	
				ATC Ile		Gly										Ph		288	•
,	IGA Arg	CAG Gln	GAT Asp	GIG Val 100	ATA Ile	gac Abp	CIG	GLn GLn	CCA Pro 105	aag Lyb	GCT Ala	GTA Val	Val	I Ile	Leu	GC:	T 3	336	i
20 G	GT Sly	ACC Thr	AAT Asn 115	GAC Asp	GTA Val	GCT Als	CAA Gln	AAT ASD 120	ACC	GGG Gly	CCG Pro	ATG Het	ACC The 125	: Ile	Glu Glu	Glu	.	384	
	3er			AAC									Gln					432	
1				GIT Val													•	480	
				CTT Lev														528	
				TAT Tyr 180														576	
		Ala		ACC Thr			Gly											624	
40											-								
	QJ As	3P G	GT (FTG C	AT C Lis P	CA A	ran i	37A (/al) !15	Ma (31y	TAT Tyr	G) n	GTG Val 220	ATG Met	GAA .	AAC Asn	ATT Ile		672
45	G1 Va 22	II L	TA C eu F	ro V	TC A	1e 3	er S	er G	IAG 1	rre (Leu)	Na !	AAG (Lys 1 235	CTG Leu	aag Lys	TA				713

(2) INFORMATION POR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 978 NOCLEOTIDES (B) TYPE: NUCLEIC ACID

 - (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

50

55

(ii) MOLECULE TYPE: GENOMIC DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

45

55

5		ATG Met	: Ala	TTO Pho	Pho	C GAT	r TT	A CCI	i CT(n COTA	10 (1) (1)	Lei	AAC Lys	Ly	A TA	CG Arg	r CCA Pro	48
		Glu	. Arç	Ty1	G10 20	ı Glı	ı Lya	a Asp) Phe	25 25	o Glu	Phe	TX	Gli	1 Gli 31	Thr	CTC Leu	96
10 [.]		GCA Ala	Glu	S AGO Ser 35	. Gyn	Lys	Phe	Pro	Leu 40	ı Asp	Pro	Val	Phe	Glu Glu 45	Arg	ATG Met	Glu	144
		TCT Ser	CAC His	Leu	Lya	ACA Thr	Val	GAA Glu 55	: Ala	TY	GAT Asp	Val	The 60	Phe	Sex	Gly	TAC	192
15		AGG Arg 65	Gly	Gln	AGG Arg	ATC Ile	Lys 70	Gly	Trp	Leu	Leu	G11 Val 75	Pro	Lys	Leu	GAA Glu	GAA Glu 80	240
20		GAA Glu	Lys	Leu	Pro	TGC Cys 85	Val	GTG Val	G) u	TAC	ATA Ile 90	Gly	TAC	AAC	Gly	GGA Gly 95	AGA Arg	288
20		gly	Phe	Pro	His 100	Asp	Trp	Leu	Phe	Trp 105	Pro	Ser	Het	Gly	177 110	Ile	Сув	. 336
25		Phe	Val	Met 115	Авр	Thr	Arg	Gly	Gln 120	Gly	Ser	Gly	Trp	Leu 125	Lys	Gly	Asp	384
		Thr	P ro 130	_	IYI	Pro	Glu	Gly 135	Pro	Val	Авр	Pro	Gln 140	Tyr	Pro	Gly	Phe	432
30		ATG Met 145	Thr	Arg	Gly	Ile	Leu 150	Asp	PTO	Arg	Thr	Tyr 155	Tyx	Tyr	Arg	Arg	Val 160	480
		TTC Phe	Thr	Asp	Ala	Val 165	Arg	Ala	Val	Glu	Ala 170	Ala	Ala	Ser	Phe	Pro 175	Gln	528
35		GTA Val	A sp	Gln	Glu 180	Arg	Ile	Val	Ile	Ala 185	Gly	Gly	Ser	Gln	Gly 190	Gly	Gly	576
	(ATA Ile	Ala	Leu 195	Ala	Val	Ser	Ala	Leu 200	Ser	Lys	Lys .	Ala	Lya 205	Alb	Leu :	Leu	624
40		TGC (Cys	GAT Asp 210	Val	Pro	Phe	Leu	TGT Cys 215	KiB	Phe	AGA . Arg .	Arg .	GCA Alb 220	GTA Val	eju Cye	Leu '	GTG Val	672

	GAT Asp 225	ACG Thr	CAT His	Pro	TAC	GCG Ala 230	Glu	ATC	ACG	AAC Asn	777 Phe 235	Leu	AAG Lys	ACC	CAC His	AGA Arg 240	720
5	gac Asp	aag Lyb	GAA Glu	GAA Glu	ATC Ile 245	OTO Val	Phe	AGG Arg	ACT	Leu 250	TCC Ser	TAT Tyr	TTC Phe	GAT Asp	GGA Gly 255	GTG Val	768
	AAC Asn	TTC Phe	GCA Alb	GCC Ala 260	AGA Arg	GCG Ala	AAG Lys	ATC Ile	CCT Pro 265	GCG Ala	CTG Leu	TTT Phe	TCT Sex	GIG Val 270	GIY	CTC Leu	816
. 10	ATG Met	gac Asp	AAC Asn 275	ATT Ile	TGT Cys	CCT Pro	CCT Pro	TCA Ser 280	ACG Thr	GTT Val	TTC Phe	GCT Ala	GCC Ala 285	TAC Tyr	aat Asn	TAC Tyr	864
15	Tyr	GCT Ala 290	GGA Gly	CCG Pro	AAG Lys	GAA Glu	ATC Ile 295	AGA Arg	ATC Ile	TAT Tyr	PTO CCG	TAC Tyr 300	aac asn	aac asn	CAC Ris	GAG Glu	864
	GGA Gly 305	GGA Gly	GCC GLY	TCT Ser	Phe	CAA Gln 310	GCG Ala	GTT Val	GAA Glu	CAG Gln	G7G Val 315	aaa Lys	TTC Phe	TTG Leu	Lys	AAA Lys 320	912
	CTA :			Lys		TAA											930

(2) INFORMATION FOR SEQ ID NO:52:

25

*3*5

40

45

50

55

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 660 NUCLEOTIDES
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: GENOMIC DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

	TTG Leu 1	Lyt	TAI	r Ph	e Ly	A GC B Ala 5	a yri	G CT	A)a	GG(G1)	, Ile	Thr	Leu	Leu	GGC Gly 15	Lev		48	
5	CTG Leu	Ali	TGG Cys	C ACC	CTC CSe O	GC Al	C TCC	GCC Ala	Gln 25	The	GAG Glu	Pro	ATC Ile	Val Val	Pbe	Val		96	
				: Se	r Gl				Авл									144	
10			3 Sei		c GG n Gly			Ser									•	192	
					C AGO L Sei		r Asr											240	•
15					C GTC Val	Arg					Asn							288	
20					AAC Asn													336	
				GJu	ACG Thr													384	
25	His		Gly		ACC			Tyr									•	432	
			-																
30					Pro								Leu					,	481
35					TCG Ser														526
					CAG Glm 180														576
40					CTG Leu			Ser											624
		æu			gjà Ggg		Arg							G					661

(2) INFORMATION FOR SEQ ID NO:53:

45

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 201 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- 55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Met Lys Val Lys His Val Ile Val Leu His Gly Leu Tyr Met Ser Gly 1 15

Leu Val Met Arg Pro Leu Cys Ser Arg Leu Glu Glu Ser Gly Val Lys 20

Val Leu Asn Leu Thr Tyr Asn Thr Arg Asp Pro Asn Arg Asp Ala Ile 35

Phe Thr Gln Ile Asp Glu Phe Ile Ser Asn Glu Pro Ser Ala Leu Val 50

Cys His Ser Met Gly Gly Leu Val Ala Arg Ala Tyr Leu Glu Ala Asn 65

Ser Ala Pro Ser His His Val Glu Lys Val Ile Thr Leu Gly Thr Pro 85

His Thr Gly Ser His Ile Ala Glu Lys Met Gln Gln Lys Gly Phe Glu 105

Leu Leu Leu Lys Asn Ser Val Glu Phe Leu Leu Ser Lys Asn Gly Asp 120

Trp Pro Phe Lys Ala Lys Leu Tyr Ser Ile Ala Gly Asp Leu Pro Ile 130

Gly Leu Met Pro Leu Ile Val Lys Gly Ser Arg Ser Asp Gly Thr Val 145

Leu Leu Asp Glu Thr Lys Leu Lys Gly Met Ala Glu His Lys Val Phe 165

His Leu Ser Ris Thr Ser Met Ile Tyr Ser Arg Gln Val Val Asn Tyr 185

Ile Leu Glu Arg Leu Asn Glu Asp Ile

- (2) INFORMATION FOR SEQ ID NO:54:
 - (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 259 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR

35

40

50

- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Me	t II	e L	/8 A:	n Ph	е Ав	p Ar	g Gl	u As	n Se	r Se	r Le	u Va	l le	u Se	r Gl
	1				5		•		1	0				1	.5
G1	A 61	y Al	la La	u Gl	y Il	e Al	a Hi	s Le	u 01 5	y Va	l Le	u Hi		p le 0.	n GJ
Ly	9 G1	я Ас 2	n Il	e Va	l Pro) AB	n Glu	110)	e Va	l Gly	Th	r Se	r Me S	t Gl	y G1)
Ile	11 5	e G1	y Al	a Se	r Met	: Ala 5:	a Ile 5	g G1	y Mei	t Lys	61 60		s Gl	u Il	e Lev
Glu 65	Gl:	u Il	e Ly	s Ası	n Phe	sei)	r Asn	Va)	l Phe	A øn 75		Ile	e Lys	s Pho	e Ser 80
Phe	: Se	. G1	y Asi	a Se: 85	Val	Val	Asp	Asr	3 Glu 90		Ile	Ala	Lys	71e 95	Phe
Asp	Thi	Le	1 Pho 100	Lys	Asp	Arg	Lys	Met 105		Asp	Thr	Va]	Ile 110		Leu
Lys	Leu	11e	Ala i	Thr	Asn	Leu	His 120	Aen	Gly	Ris	Lys	Lys 125		Phe	The
Ala	9er 130	Asp	qeA ı	Val	Leu	Ile 135	Lya	Asp	Ala	lle	140	Ser	Thr	Met	Ala
Ile 145	Pro	Gly	Val	Phe	Glu 150	ßЛп	His	Ile	Ile	Asp 155	Gly	Glu	Thr	Tyr	Gly 160
Двр	Gly	Phe	Leu	Cys 165	Glu	Asn	Leu	gly	Val 170	λen	Glu	Ala	The	Phe 175	Asn
Авр	Val	Leu	Ala 180	Val	Авр	Val	Het	01y 185	Glu	Asn	9er	Phe	Glu 190	Lys	Ala
Met	Pro	A8p 195	Asn	Phe	Phe .	Lys	Thr 200	Ser	Asn	Val	Leu	Glu 205	Mat	Phe	Glu
Lys	Ser 210	Met	Arg	Leu	Phe	11e 215	Tyr	Asn	Gln		Gln 220	Thr	His	Ile	Lys
Asn 225	Ala	Asn	Lys	Asn	lle 230	Tyr	Leu	Ila	Glu	Pro 235	Val	Thr	Lys	Glu	Tyr 240
Lys	Thr	Phe	Gln	Phe 245	His	Lya	His !	Lys	Glu 250	Ile :	Arg	Ala	Leu	Gly 255	Leu
31v	Leu	Leu													

(2) INFORMATION FOR SEQ ID NO:55:

10

15

20

25

30

35

45

50

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 301 AMINO ACIDS
 - (B) TYPE: AMINO ACID-
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Met Pro Leu His Pro Lys Val Lys Leu Leu Ser Gln Leu Pro Pro 15

Gln Asp Phe Ser Arg Asn Val Gln Asp Leu Arg Lys Ala Trp Asp Leu 20

Pro Phe Ser Gly Arg Arg Glu Thr Leu Lys Arg Val Glu Asp Leu Glu 15

The Pro Thr Arg Asp Ala Arg Ile Arg Ala Arg Val Tyr Thr Pro Ser 50

- Artista

(2) INFORMATION FOR SEQ ID NO:56:

15

20

25

30

35

45

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 326 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Val Ala Phe Phe Asp Mat Pro Leu Glu Glu Leu Lys Lys Tyr Arg Pro 15

Glu Arg Tyr Glu Glu Lys Asp Phe Asp Glu Phe Trp Arg Glu Thr Leu Lys Glu Ser Glu Gly Phe Pro Leu Asp Pro Val Phe Glu Lys Val Asp Phe His Leu Lys Thr Val Glu Thr Tyr Asp Val Thr Phe Ser Gly Tyr So Gly Gln Arg Lle Lys Gly Trp Leu Leu Val Pro Lys Leu Ala Glu Glu Lys Leu Pro Cys Val Val Gln Tyr Ile Gly Tyr Asn Gly Gly Arg 90

Gly Phe Pro His Asp Trp Leu Phe Trp Pro Ser Met Gly Tyr Ile Cys

Phe Val Met Asp Thr Arg Gly Gln Gly Ser Gly Trp Met Lys Gly Asp 115

Thr Pro Asp Tyr Pro Glu Gly Pro Val Asp Pro Gln Tyr Pro Gly Phe 130

Met Thr Arg Gly Ile Leu Asp Pro Gly Thr Tyr Tyr Tyr Arg Arg Val 145

Met Thr Arg Gly Ile Leu Asp Pro Gly Thr Tyr Tyr Tyr Arg Arg Val 145

Phe Val Asp Ala Val Arg Ala Val Glu Ala Ala Ile Ser Phe Pro Arg 165

Val Asp Ser Arg Lys Val Val Val Ala Gly Gly Ser Gln Gly Gly Gly 180

Ile Pro Leu Ala Val Ser Ala Leu Ser Asn Arg Val Lys Ala Leu Leu 195

Cys Asp Val Pro Phe Leu Cys His Phe Arg Arg Ala Val Gln Leu Val 210

Asp Thr His Pro Tyr Val Glu Ile Thr Asn Phe Leu Lys Thr His Arg 225

Asp Lys Glu Glu Ile Val Phe Arg Thr Leu Ser Tyr Phe Asp Gly Val 245

Asn Phe Ala Ala Arg Ala Lys Val Pro Ala Leu Phe Ser Val Gly Leu 260

Met Asp Thr Ile Cys Pro Pro Ser Thr Val Phe Ala Ala Tyr Asn His 290

Gly Gly Gly Ser Phe Gln Ala Ile Glu Gln Val Lys Phe Leu Lys Arg 305

Cly Gly Gly Gly Ser Phe Gln Ala Ile Glu Gln Val Lys Phe Leu Lys Arg 305

Leu Phe Glu Glu Glu Glu Ile

(2) INFORMATION FOR SEQ ID NO:57:

15

20

25

30

35

45

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 298 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Met Arg Thr Leu Ser Phe Gly Pro Met Thr Thr Gly Gly Ser Ile His 15 Met Ala Thr Met Asp Val Met Arg Gly Pro Gly Met Gln Arg Leu Ser 20 Gln Gly Ala Arg Glu Ala Ala Asn His Pro Trp Ala Lys Arg Leu Gly Arg Met Gly Tyr Ala Ala Lys Gly Ala Val Tyr Ala Ile Ile Gly Val 55 Gly Ala Val Tyr Ala Ile Ile Gly Val 60 His Gly Ala Leu Lys Leu Ala Ala Gly Glu Gly Gly Arg Thr Thr Asp Ser 75 Re 80 His Gly Ala Val Leu Val Asn Thr Val Ala His Gly Pro Phe Gly Val Ala Leu 95 Leu Ala Val Leu Val Val Gly Leu Leu Gly Tyr Val Val Trp Arg Phe 100 Ala Gln Ala Phe Val Asp Thr Glu Asp Lys Gly Ser Asp Ala Lys Gly 115 Pro Phe Val Asp Thr Glu Asp Lys Gly Ser Asp Ala Lys Gly 115 Pro Phe Val Asp Thr Glu Asp Lys Gly Ser Asp Ala Lys Gly

20

30

35

10

15

Ile Ala Thr Arg Ala Met Tyr Phe Leu Ser Gly Cys Ile Tyr Ala Ser 130 Thr Arg Ala Met Tyr Phe Leu Ser Gly Cys Ile Tyr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ala His Gly Arg 165 Ser Lys Gly Thr Gln Gly Trp Thr Ala Thr Leu Met Glu Gln Pro Phe 165 Gly Arg Val Leu Val Ala Leu Val Gly Leu Gly Ile Val Gly Phe Ala 180 Leu Lys Gln Phe Ris Thr Ala Trp Lys Ala Lys Phe Arg Glu Lys Leu Inst Thr Gly Leu Ala Ala Arg Lys Gln His His Ile Glu Arg Met 210 Thr Gly Leu Ala Ala Arg Lys Gln His His Ile Glu Arg Met 220 Cys Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly 230 Gly Phe Leu Val Arg Ser Ala Val Asp Ala Asn Pro Gly Glu Ala Lys 255 Gly Leu Gly Glu Ala Leu Ala Val Val Ala Arg Gln Pro Ser Gly Asp 260 Leu Leu Gly Val Val Ala Arg Gly Leu Val Ala Tyr Ala Ala Tyr 275 Leu Phe Leu Gln Ala Arg Tyr Arg Glu Leu Val Ala Tyr Ala Ala Tyr 290 Leu Phe Leu Gln Ala Arg Tyr Arg Glu Leu Phe Leu Gln Ala Arg Tyr Arg Glu Leu

45

50

55

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 304 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Met Ser Lys Phe Ala Ile Leu Trp Ala Leu Ile Thr Ala Tyr Leu Pro 15

Glu Pro Val Met Lys Leu Val Tyr Leu Gly Arg Arg Glu Thr Leu Gly 20

Ala Arg Thr Leu Asp Val Lys Ala Gln Ala Val Gly Arg Leu Ala Asm 45

Ala Thr Arg Pro Val Gly Val Ile Pro Thr Val Glu Glu Ser Arg Lys 50

Met Thr Asp Lys Ala Val Ser Leu Phe Asp Gln Pro Ala Pro Glu Leu 65

Phe Arg Lys Lys Asp Ile Gln Ile Asp Gly Ala Glu Gly Pro Ile Asp 90

Ala Arg Ile Tyr Ser Gly Pro Ala Lys His Arg Pro Arg Pro Ile Leu 100

Val Tyr Phe His Gly Gly Gly Trp Val Gln Gly Asn Leu Asp Ser His 135

Asp Gly Val Cys Gly Lys Leu Ala Lys Trp Ala Asn Cys Ile Val Ile 130

Ser Val Asp Tyr Arg Leu Ala Pro Glu His Lys Phe Pro Cys Ala Pro 145

Leu Asp Ala Ile Ala Ala Tyr Lys Trp Val Arg Ala Asn Ala Thr Asn 175

25

30

45

50

15

Leu Gly Gly Asp Pro Glu Arg Ile Gly Val Gly Asp Ser Ala Gly 180

Gly Asn Leu Ala Ala Val Val Cys Gln Gln Thr Ala Met Asn Gly Glu 200

Arg Thr Pro Asp Leu Gln Val Leu Ile Tyr Pro Ala Leu Asp Ala Arg 210

Met Ile Ser Thr Ser Met Glu Glu Leu Arg Asp Ala Tyr Ile Leu Pro 225

Lys Ser Arg Met Glu Tyr Phe Leu Gly Leu Tyr Thr Arg Gly Pro Asp 245

Asp Ile Glu Asp Leu Arg Met Ser Pro Ile Leu Arg Asp Thr Val Ala 260

Asp Gln Pro Gln Ala Cys Ile Val Thr Cys Gly Phe Asp Pro Ala Arg 275

Arg Arg Glu His Leu Arg Arg Thr Leu Asn Cys Arg Gly Asp Arg Arg Arg 290

- (2) INFORMATION FOR SEQ ID NO:59:
 - (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 308 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

30

25

10

Ser Pro Leu Gln Ala Asp Phe Ser Gly Met Pro Pro Val Phe Leu Thr 235

Ala Ser Asp Ser Glu Ile Leu Leu Asp Asp Cys Leu Arg Met Ala Asp 255

Ris Leu Arg Ala Gln Gly Val Val Val Thr Asp Arg Ile Val Glu Asn 265

Ris Pro His Val Trp His Ile Phe Gln Arg Leu Leu Pro Glu Ala Asp 275

Gln Gly Leu Arg Ala Ile Ala Ala Trp Ile Lys Pro Leu Leu Ser Gly 290

Ser Asn Glu Ser 305

45

50

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 237 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Met Leu Thr Phe Asn Val Leu Tyr Gly Met Net Lys Gln Lys Leu Ala 15

Ala Ile Leu Met Phe Leu Gly Leu Ser Ala Ala Glu Ala Gln Asp Trp 25

Pro Asp Leu Gln Lys Tyr Arg Ser Ala Asn Lys Glu Ala Leu Leu Leu Gly Leu Son Asn Lys Glu Ala Lys Leu Leu Gly Lys Glu Asn Arg Lys Val Val Phe Met Gly Asn Ser Ile Thr Glu Gon Ala Trp Ile Ser Gln Arg Pro Glu Phe Phe Ser Glu Asn Gly Phe Ile Gon Arg Gly Arg Gly Ile Ser Gly Gln Thr Thr Pro Gln Met Leu Leu Arg Phe 95

Arg Gln Asp Val Ile Asp Leu Gln Pro Lys Ala Val Val Ila Leu Ala 100

Gly Thr Asn Asp Val Ala Gln Asn Thr Gly Pro Met Thr Ile Glu Glu 115

Ser Leu Ala Asn Ile Lys Ser Met Val Glu Leu Ala Gln Asn Gly 130

Ile Thr Pro Val Leu Cys Thr Val Leu Pro Ala Asp Arg Phe Ser Trp 165

Arg Pro Glu Leu Thr Pro Ala Glu Thr Ile Ile Ala Leu Asn Gln Leu 165

Ile Lys Gln Tyr Ala Glu Ala Gln Gly Leu Ala Leu Val Asp Tyr Ris 180

Ala Ala Leu Thr Asn Lys Gly Gly Gly Leu Pro Val Lys Tyr Gly Glu Asp Cly Val His Pro Asn Val Ala Gly Tyr Gln Val Met Glu Asn Ile 210

Val Leu Pro Val Ile Ser Ser Glu Leu Ala Lys Leu Lys 235

- (2) INFORMATION FOR SEQ ID NO:61:
- (2) INFORMATION FOR SEQ ID NO:61:

10

15

20

25

30

35

- (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 326 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Me	t 1	Ala	Ph	e Pì	e As	p Le S	u Pr	o Lei	u Gl	u G1 1	u Le O	u Ly	8 Ly	в Ту		g Pr
G)	u i	Arg	Ty	* G)	u G1	u Ly:	a Asj	p Pho	e Aø _j 2:	p G 1	u Ph	e Tx	p (31)	u Gl		r Len
Al	a (31u	Se:	r G1 5	u Ly	s Ph	e Pro	Let 40	AS ₁	Pr	o Vai	l Pb	e Gl:		g Me	t GJ/
Se	r J	ais 50	Le	u Ly	e Th	r Val	1 Glu 55	ı Ala i	Туз	. Yal	y Val	Thi 60		e Se	r Gl	y Tyz
Ar		lly	Gli	n Ar	g Il	e Lys	Gly	/ Trp	Leu	Let	1 Va] 75		Ly	Let	. Gl	4 Glu 80
Gli	ı L	ya	Lev	Pr	65 85	a Val	. Val	. Gln	Tyr	11e	Gly	בעד י	Asr	g GL	7 G1 ₃	y Arg
61)	P	he	Pro	Hi:	Asp)	Trp	Leu	Phe	Trp 105	Pro	9er	Met	Gly	110	Ile	Cys
Phe	V	al	Met 115	Yel	Thr	Arg	Gly	Gln 120	Gly	Ser	Gly	Trp	Leu 125	Lув	Gly	Азр
Thr		10 30	Двр	Tyz	Pro	Glu	Gly 135	Pro	Val	Asp	Pro	01n 140		Pro	Gly	Phe
Met 145	T	hr	Arg	Gly	Ile	Leu 150	Авр	Pro	Arg	Thr	Tyr 155	Tyr	Tyr	Arg	Arg	Val 160
Phe	Tì).	Авр	λla	Val 165	Arg	Ala	Val	Glu	Ala 170	Ala	Ala	Ser	Phe	Pro 175	Gln
Val	Α£	3p	Gln	Glu 180	Arg	Ile	Val	Ile	Ala 185	Gly	Gly	Ser	Gln	Gly 190	Gly	Gly
Ile	A	la .	Leu 195	Ala	Val	Ser	Ala	Leu 200	Ser	Lys	Lys	Ala	Lys 205	Ala	Leu	Leu
Сув	A:	D D	Val	Pro	Phe	Leu	Cys 215	His	Phe	Arg	λrg	Ala 220	Val	Gln	Leu	· Val
Д8р 225	Tì	ır i	His	Pro	Tyr	Ala 230	Glu	Ile	Thr	Asd	Phe 235	Leu	Lys	Thr	His	Arg 240
ASP	L)	/8 (Glu	Glu	11e 245	Val	Phe	Arg	Thr	Leu 250	Ser	Tyr	Phe	λвр	Gly 255	Val
Asn	Ph	e .	Ala	Ala 260	Arg	Ala	Lys	Ile	Pro 265	Ala	Leu	Phe	Ser	Val 270	Gly	Leu
Met	As		Aan 275	Ile	Cys	Pro	Pro	Ser 280	Thr	Val	Phe	Ala	Ala 285	Tyr	Asn	Tyr
Tyr	29		31y	Pro	Lys		11e 295	Arg	Ile	Tyr	Pro	Tyr 300	Asn	Asn	His	Glu
Gly 305	Gl	у (ily	Ser	Phe	Gln 310	Ala	Val	Glu		Val 315	Lys	Phe	Leu	Lys	Lys 320
Leu	Ph	e (lu	Lys	Gly 325											

- (2) INFORMATION FOR SEQ ID NO:62:
 - (i) SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 220 AMINO ACIDS
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR

(11) MOLECULE TYPE: PROTEIN

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Leu Lys Tyr Phe Lys Ala Arg Leu Ala Gly Ile Thr Leu Leu Gly Leu 1

Leu Ala Cys Thr Ser Ala Ser Ala Gln Thr Glu Pro Ile Val Phe Val 20

His Gly Tyr Ser Gly Ser Ala Ser Asn Trp Asp Thr Met Leu Gly Arg 45

Phe Arg Ser Asn Gly Tyr Ala Ser Gly Ser Leu Tyr Thr Phe Asn Tyr 50

Asn Ser Leu Val Ser Ser Asn Arg Thr Ser Ala Ser Glu Leu Arg Ser 65

Phe Val Asn Thr Val Arg Ser Arg His Gly Asn Ala Arg Ile Ala Leu 85

Phe Val Asn Thr Val Arg Ser Arg His Gly Asn Ala Arg Ile Ala Leu 90

Val Ala His Ser Asn Gly Gly Leu Val Ser Arg Trp Tyr Arg Ala Glu 100

Leu Gly Gly Glu Thr Ala Thr Arg Arg Phe Val Thr Leu Gly Thr Pro 125

His Arg Gly Thr Thr Trp Ala Tyr Ala Cys Tyr Ser Pro Ala Cys Phe 130

Glu Met Arg Pro Gly Ser Ser Leu Leu Thr Thr Leu Gly Ser Arg Ala 145

Cys Asp Arg Ser Leu Trp Ser Asn Thr Asp Gly Ile Ile Leu Pro Ala 165

Ser Ser Ala Gln Cys Gly Val Ser Thr Arg Thr Ala Asp Val Ser His 180

Leu Asp Leu Leu Thr Asp Ser Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gln End Gly End Arg Cys Thr Glu Arg Ala Pro 6

Leu Gly Fire Thr Glu Arg Cys Thr Glu Arg Ala Pro 6

Leu Gly F

35

40

50

55

10

15

20

25

SEQUENCE LISTING

[0133]

- (1) GENERAL INFORMATION:
 - (i) APPLICANT: Diversa Corporation
 - (ii) TITLE OF INVENTION: ESTERASES
 - (iii) NUMBER OF SEQUENCES: 62
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Morrison & Foerster LLP
 - (B) STREET: 3811 Valley Centre Drive, Suite 500
 - (C) CITY: San Diego
 - (D) STATE: CA
 - (E) COUNTRY: USA
- (F) ZIP: 92130
 - (v) COMPUTER READABLE FORM:

	(A) MEDIUM TYPE: Diskette	
	(B) COMPUTER: IBM Compatible	
	(C) OPERATING SYSTEM: Windows 95	
	(D) SOFTWARE: FastSEQ for Windows Version 2.0b	
5		
	(vi) CURRENT APPLICATION DATA:	
	(A) APPLICATION NUMBER: US97/02039	
	(B) FILING DATE: 11-FEB-1997	
10	• •	
	(vii) PRIOR APPLICATION DATA:	•
	(A) APPLICATION NUMBER: 08/602,359	
	(B) FILING DATE: 16-FEB-1996	
15	(5) (12.11.2.5.11.2.5.12.5.10.5.	
	(viii) ATTORNEY/AGENT INFORMATION:	
	(A) NAME: Einhorn, Ph.D., D.D.S., Gregory	
	(B) REGISTRATION NUMBER: 38,440	
20	(C) REFERENCE/DOCKET NUMBER: 09010/010WO1	
	(O) THE ETENOE BOOKET NOMBER. 09010/010WOT	
	(ix) TELECOMMUNICATION INFORMATION:	
	(A) TELEPHONE: 858 720 5133	
25	(B) TELEFAX: 858 720 5125	
	(2) INFORMATION FOR SEQ ID NO:1:	
	(i) SEQUENCE CHARACTERISTICS:	
30		
	(A) LENGTH: 52 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
35	· ,	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:	
40		
40	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGTCTTTA AACAAGCACT CT	6.0
	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGTCTTTA AACAAGCACT CT	52
	(2) INFORMATION FOR SEQ ID NO:2:	
	(I) O O O O O O O O O O O O O O O O O O O	
45	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 31 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
50	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:	
55	IN OF GOTHER POOL OF A ID MO.2.	
-		
	CGGAAGATCT CTATCGTTTA GTGTATGATT T	31

	(2) INFORMATION FOR SEQ ID NO:3:	
	(i) SEQUENCE CHARACTERISTICS:	
5	(A) LENGTH: 52 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
10	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:	
15	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGAAACTC CTTGAGCCCA CA	52
	(2) INFORMATION FOR SEQ ID NO:4:	
20	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 31 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
25	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:	
30	CGGAAGATCT CGCCGGTACA CCATCAGCCA C	31
	(2) INFORMATION FOR SEQ ID NO:5:	
35	(i) SEQUENCE CHARACTERISTICS:	
40	(A) LENGTH: 52 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
40	(ii) MOLECULE TYPE: cDNA	
45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:	
	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCCATAT GTTAGGAATG GT	52
50	(2) INFORMATION FOR SEQ ID NO:6:	
	(i) SEQUENCE CHARACTERISTICS:	
55	(A) LENGTH: 53 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	

	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:	
5	CGGAGGTACC TTAGAACTGT GCTGAAGAAA TAAATTCGTC CATTGCTCTA TTA	53
	(2) INFORMATION FOR SEQ ID NO:7:	
10	(i) SEQUENCE CHARACTERISTICS:	
15	(A) LENGTH: 49 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:	
	CGGAGGTACC TTAGAACTGT GCTGAAGAAA TAAATTCGTC CATTGCTCT	49
25	(2) INFORMATION FOR SEQ ID NO:8:	
23	(i) SEQUENCE CHARACTERISTICS:	
<i>30</i>	(A) LENGTH: 53 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:	
	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGAGATTG AGGAAATTTG AAG	53
40	(2) INFORMATION FOR SEQ ID NO:9:	
	(i) SEQUENCE CHARACTERISTICS:	
45	(A) LENGTH: 31 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
50	(ii) MOLECULE TYPE: cDNA	
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:	
55	CGGAGGTACC CTATTCAGAA AGTACCTCTA A	31
- -	(2) INFORMATION FOR SEQ ID NO:10:	
	(i) SECUENCE CHARACTERISTICS:	

	(A) LENGTH: 52 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
5	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:	
10	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGTTTAAT ATCAATGTCT TT	52
	(2) INFORMATION FOR SEQ ID NO:11:	
15	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 31 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
	CGGAAGATCT TTAAGGATTT TCCCTGGGTA G	31
30	(2) INFORMATION FOR SEQ ID NO:12:	
	(i) SEQUENCE CHARACTERISTICS:	
35	(A) LENGTH: 52 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:	
	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGGAGGTT TACAAGGCCA AA	52
45	(2) INFORMATION FOR SEQ ID NO:13:	
	(i) SEQUENCE CHARACTERISTICS:	
50	(A) LENGTH: 31 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
55	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:	

	CGGAGGTACC TTATTGAGCC GAAGAGTACG A	31
_	(2) INFORMATION FOR SEQ ID NO:14:	
5	(i) SEQUENCE CHARACTERISTICS:	
10	(A) LENGTH: 53 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGATTGGC AATTTGAAAT TGA	53
20	(2) INFORMATION FOR SEQ ID NO:15:	
	(i) SEQUENCE CHARACTERISTICS:	
25	(A) LENGTH: 31 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
30	(ii) MOLECULE TYPE: cDNA	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:	
35	CGGAGGTACC TTAAAGTGCT CTCATATCCC C	31
	(2) INFORMATION FOR SEQ ID NO:16:	
	(i) SEQUENCE CHARACTERISTICS:	
40	(A) LENGTH: 52 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
45	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:	
50	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCCAGCT AATGACTCAC CC	52
	(2) INFORMATION FOR SEQ ID NO:17:	
55	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 32 base pairs	
	(B) TYPE: nucleic acid	

	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:	
	CGGAAGATCT TCAACAGGCT CCAAATAATT TC	32
10	(2) INFORMATION FOR SEQ ID NO:18:	
	(i) SEQUENCE CHARACTERISTICS:	
15	(A) LENGTH: 29 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
20	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:	
25	CGGAAGATCT ACAGGCTCCA AATAATTTC	29
	(2) INFORMATION FOR SEQ ID NO:19:	
30	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 52 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
35	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:	
40	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCTTGAT ATGCCAATCG AC	52
45	(2) INFORMATION FOR SEQ ID NO:20:	
45	(i) SEQUENCE CHARACTERISTICS:	
50	(A) LENGTH: 31 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
55	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:	
	CGGAGGTACC CTAGTCGAAC AGAAGAAGAG C	31

	(2) INFORMATION FOR SEQ ID NO:21:	
	(i) SEQUENCE CHARACTERISTICS:	
5	(A) LENGTH: 52 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
10	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:	
15	CCGAGAATTC ATTAAAGAGG AGAAATTAAC TATGCCCCTA GATCCTAGAA TT	52
	(2) INFORMATION FOR SEQ ID NO:22:	
20	(i) SEQUENCE CHARACTERISTICS:	
20	(A) I ENGTH: 21 base pairs	
	(A) LENGTH: 31 base pairs (B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
25		
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
30		
	CGGAGGTACC TTAAATTTTA TCATAAAATA C	31
	(2) INFORMATION FOR SEQ ID NO:23:	
35		
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 555 base pairs	
	(B) TYPE: nucleic acid	
40	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: Genomic DNA	
45	(ix) FEATURE:	
	(A) NAME/KEY: Coding Sequence	
	(B) LOCATION: 1552	
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	

						His					T ATO P Met O						48	
					Pro					e Al	A TTI a Lei			e Se			96	
10				Asn					n Al		A AGO			u G			144	
15			Ser					Ile			T CT! n Let		y Le				192	
		Ile					Val				A AA1 g Asi 75	ı Lei				AGT Ser 80	240	
20																		
25											GGT Gly				n L		288	
. 5											ATA Ile		Phe				336	
30							Ile				TTC Phe	Ile					384	
35											ATA Ile						432	
					His						CCG Pro 155				a I)		480	
40				Leu .							CCA				y As		528	
45			AAA Lys						TAG								555	
	(2)	INFOE	·	ON F	OR SI	EO ID	NO-2	4 ·										

50

*5*5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1041 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA

								•								
	(ix)	FEAT	TURE:					٠								
5	(A) NAME/KEY: Coding Sequence (B) LOCATION: 11038															
	(xi)	SEQ	UENC	E DE	SCRIF	PTION	I: SEC) ID N	O:24:							
10	ATG Met 1	AAA Lys	CTC Leu	CTT Leu	GAG Glu 5	CCC Pro	ACÁ Thr	AAT Asn	ACC Thr	TCC Ser 10	TAC Tyr	ACG Thr	CTG Leu	TTA Leu	CAG Gln 15	GAT Asp
	TTA	GCA	TTG	CAT	TTT	GCA	TTT	TAC	TGG	TTT	CTG	GCC	GTG	TAT	ACG	TGG
15																
20																
25																
30																
35												á				

	Leu	Ala	Leu	His 20	Ala	Phe	Tyr	Trp 25	Phe	Leu	Ala	Val	Tyr 30	Thr	Trp		
5					GTC Val												144
10					GGC Gly												192
15					CTT Leu 70												240
					CAC His												288
20					GAG Glu												336
25					CTG Leu												384
30					TAT Tyr											,	432
					CTC Leu 150											•	480
35					GGG Gly											!	528
40					GCC Ala											!	576
45					ATA Ile											(624
					GGT Gly											(672
					TAT Tyr 230											•	720

5						Leu					Gly			TTC - Phe	768
•					Met									GCG	816
10				Val		AAG Lys			Leu						. 864
15						GAC Asp									912
20						GAG Glu 310									960
						GGC Gly									1008
25						CCG Pro					TGA				1041
30	(2)					EQ ID									
35		((A) LE B) TY C) ST	NGTH PE: ni	: 789 ucleic EDNI	ESS: s	pairs	.							
40			OLEC		TYPE:	: Geno	mic D	NA							
45		(B) LO	CATIC)N: 1.	oding 786 CRIPT	•		D NO:	25:					
50						AAT Asn									48
5 5						CCA Pro									96
						GAG Gl u									144

													•	
			35		,			40			45			
5									CAT His					192
10			-						AAC Asn					240
									AAA Lys					288
15									TAC Tyr 105					336
20									GGT Gly					384
25									AAG Lys					432
20					-				GAT Asp					480
30									AAA Lys					528
35									AGA Arg 185					576
									TGT Cys					624
40						Leu			AGC Ser					672
45					Ser				ATC Ile					720
50									GTT Val					768
	ATT	TCT	TCA	GCA	CAG	TTC	TAA							789

Ile Ser Ser Ala Gln Phe 260

	(2) INF	ORM	ATION	FOR	SEQ	ID NO	D:26:								
10	(i)	SEQL	JENCI	E CHA	ARAC'	TERIS	STICS	:							
		(B) (C)	TYPE: STRA	nucle NDED	56 bas eic aci ONESS f: linea	d S: sinç									
15	(ii)	MOL	ECUL	E TYF	E: Ge	enomi	c DNA	١							•
•	• •		TURE												
20	-	(A) I	NAME	KEY:	Codii 175	_	quenc	е			•				
,	(xi) SEC	UENC	DE DE	SCRI	PTIOI	N: SE	Q ID N	10:26	:					
25															
								GAG Glu							48
30								ATA Ile							96
35								TTA Leu 40							144
	GTT Val							Gly							192
40								CTG Leu							240
45								GAG Glu						_	288
50								GAA Glu							336
								GGA Gly 120							384
55								GGC Gly							432

5 ·		Pro			TAT							Asp			480
					GTT Val 165										528
10														ATT Ile	576
15					TTG Leu										624
20					GAG Glu										672
					CTT Leu										720
25					GCC Ala 245							TAG			756
30	(2) IN	FORM	IATIO	N FO	R SEC	ID N	0:27:								
	(i)	SEQ	UENC	E CH	ARAC	TERI	STICS	S:							
35		(B) (C)	TYPE STRA	: nucl	394 ba leic ac DNES Y: line	id S: sin									
40	(ii) MOL	-ECUI	E TY	PE: G	enomi	ic DN	A							
	(i:	x) FEA													
45					/: Cod l: 18	-	quen	ce			-				
	(>	d) SEC	QUEN	CE D	ESCR	IPTIO	N: SE	Q ID I	NO:27	' :					
50					AAT Asn 5										48
				Met	AAG Lys				Glu				Glu		96
55				20					25				30		

5	GTA 7									144
. 10	TTG I									192
	AGT (Ser \ 65									240
15	GAT (288
20	GTG (336
25	TAT (384
25	AGT A									432
30	ATC (Ile A 145									480
35	ACC C									528
40	CAT T				Arg				-	576
40	TTT (Gln								624
45	AAG C Lys I 2									672
50	GCA G Ala G 225									720
•	AAT C									768

					245					250	•					25	5		
5	TCA Ser																		816
10	TTT Phe												sp]						864
	GTT /									TAA	١								894
15	(2) INF	ORM	IATIOI	N FOI	R SEC	ID N	D:28:												
	(i)	SEQ	UENC	E CH	ARAC	TERIS	STICS	S :											
20		(B) (C)	TYPE STRA	: nucl	789 ba leic ac DNES Y: line	id S: sin													
25	, ,) FEA	TURE	: :	PE: G														
					/: Cod l: 17		quen	ce											
30	(xi) SEC	QUEN	CE D	ESCR	IPTIO	N: SE	Q ID	NO:2	8:									
35		Gli			C AAG					y G									48
					T GG														96
40				2	0	-			2	25		_	-	-	3	30	_		
				ı Le	C AAG u Asi			a Gl							r Pi				144
45			y His		G AA			o G1						Th					192
50		Ala			A ATO		a As												240
					C GGG														288
	PIO	rii6	= LE	u PN	8.	_		. ne	u GI	_	90	₽€ſ	1111	vd	. 11	1	95	TAT	
55	GCT	GAG	G AC	G CG	G CC	C GA	r aa	TA A	A CG	G G	GA	TTA	ATA	GC	т то	cc 1	rcg	CCT	336

	Ala	Glu	Thr	Arg 100	Pro	Asp	Lys	·Ile	Arg 105	GIA	Leu	Ile	Ala	Ser 110	Ser	Pro		
5															CTC Leu	GCG Ala		384
10 .															GGC			432
15															TAC			480
-															AGG Arg 175			528
20													Asp		ATA Ile			576
25															CCG Pro			624
30															AAA Lys			672
															CCC Pro			720
35															GAA Glu 255		-	768
40		TAC Tyr					TAA											789
	(2) INF	ORM	OITA	N FOF	SEQ	ID NO	D:29:											
45	(i)	SEQ	JENC	E CH	ARAC	TERIS	STICS	:										
		(B)	TYPE	: nucle	50 ba: eic aci ONES:	d .												
50					Y: line:	_												
	(ii)) MOL	ECUL	E TY	PE: Ge	enomi	c DNA	\										
55	(i)	() FEA	TURE	: :														
					Codi		quenc	е										

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

5			TTG Leu 5									•	48
. 10	 -		GCT Ala			_	_	_	_				96
			CTC Leu										144
15			GTT Val										192
20			CTC Leu										240
•			AAA Lys 85										288
25			GAG Glu										336
30			TGC Cys										384
		AGA	GAC Asp		CCC					AGT			432
35			GAA Glu										480
40			AAC Asn 165										528
45			G] y										576
			CTC Leu										624

	CGT	TCC	AAT	TCG	GAA	AAG	AGA	AAG	GAG	TTC	TGC	AAC	GTA	GTT	ATA	GAA	672
5	Arg	Ser 210		Ser	Glu	Lys	Arg 215	-	Glu	Phe	Cys	Asn 220		Val	Ile	Glu	
		Pro					Ser		CTG Leu			Asn					720
10					GAT Asp 245				CTT Leu	TAA							750
15	(2) INF	ORM	OITA	I FOR	SEQ	ID NO	D:30:										
	(i)	SEQU	JENC	E CHA	ARAC	TERIS	STICS	:							•		
20		(B) (C)	TYPE: STRA	nucle	017 ba eic aci ONESS 7: linea	d S: sing											
25					PE: Ge	enomi	c DNA	١									
	(IX) FEA			Codi	na Se	nuonn	۰۵									
30					110	-	quenc	.6									
	(xi) SEC	UENC	CE DE	SCRI	PTIO	N: SE	A DI D	1O:30:	!							
35									ATC Ile								48
40									ATT Ile 25								96
4E									AAG Lys								144
45									GGA								192
	Trp	Leu 50	Glu	Leu	Asp	Ala	55	Asn	Gly	Val	Thr	Leu 60	Ala	Gly	Glu	Leu	
50									TCC Ser								240
5 <i>5</i>									TGG Trp								288

	TAT GC															336
5	TTT CG		Asn													384
10	GGC AT	e Phe														432
15	GCC AT Ala Il 145															480
	TCA CT Ser Le															528
20	CTC GC Leu Al															576
25	CCC GC. Pro Al		Thr													624
30	CGC TA Arg Ty 21	r Phe														672
	GCT TT Ala Ph 225															720
35	CTT GA Leu As															768
40	TCA GTO	l Ser	Glu 260	Tyr	Phe	Lys	Ser	Tyr 265	Thr	Leu	Thr	Gly	Gln 270	Lys	Leu	816
45	GCG TT Ala Ph	e Leu 275	Asn	Cys	Pro	Ser	Tyr 280	Ile	Leu	Ala	Ala	Gly 285	Asp	Asp	Pro	864
	ATA AT Ile Il 29	e Pro														912
50	CAC AT His Il 305					Gln										960

5 ·	CTG CAT AAA CCT AGT GCT GCC GAC AAA TAT GCG GTG AAA TTA TTT GGA Leu His Lys Pro Ser Ala Ala Asp Lys Tyr Ala Val Lys Leu Phe Gly 325 330 335 GCC TGT TGA Ala Cys	1008
10		
	(2) INFORMATION FOR SEQ ID NO:31:	
	(i) SEQUENCE CHARACTERISTICS:	
15	(A) LENGTH: 936 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
20 .	(ii) MOLECULE TYPE: Genomic DNA	
	(ix) FEATURE:	
25	(A) NAME/KEY: Coding Sequence (B) LOCATION: 1933	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:	
30		
	ATG CTT GAT ATG CCA ATC GAC CCT GTT TAC TAC CAG CTT GCT GAG TAT Met Leu Asp Met Pro Ile Asp Pro Val Tyr Tyr Gln Leu Ala Glu Tyr 1 5 10 15	48
35	TTC GAC AGT CTG CCG AAG TTC GAC CAG TTT TCC TCG GCC AGA GAG TAC Phe Asp Ser Leu Pro Lys Phe Asp Gln Phe Ser Ser Ala Arg Glu Tyr 20 25 30	96
	AGG GAG GCG ATA AAT CGA ATA TAC GAG GAG AGA AAC CGG CAG CTG AGC Arg Glu Ala Ile Asn Arg Ile Tyr Glu Glu Arg Asn Arg Gln Leu Ser 35 40 45	144
40	CAG CAT GAG AGG GTT GAA AGA GTT GAG GAC AGG ACG ATT AAG GGG AGG	192
	Gln His Glu Arg Val Glu Arg Val Glu Asp Arg Thr Ile Lys Gly Arg 50 55 60	
45	AAC GGA GAC ATC AGA GTC AGA GTT TAC CAG CAG AAG CCC GAT TCC CCG Asn Gly Asp Ile Arg Val Arg Val Tyr Gln Gln Lys Pro Asp Ser Pro 65 70 75 80	. 240 .
50	GGT CTG GTT TAC TAT CAC GGT GGT GGA TTT GTG ATT TGC AGC ATC GAG Val Leu Val Tyr His Gly Gly Gly Phe Val Ile Cys Ser Ile Glu 85 90 95	288
	TCG CAC GAC GCC TTA TGC AGG AGA AYY GCG AGA CTT TCA AAC TCT ACC Ser His Asp Ala Leu Cys Arg Arg Ile Ala Arg Leu Ser Asn Ser Thr 100 105 110	336
55	GTA GTC TCC GTG GAT TAC AGG CTC GCT CCT GAG CAC AAG TTT CCC CCC Val Val Ser Val Asp Tyr Arg Leu Ala Pro Glu His Lys Phe Pro Ala	384

			115					120			125				
5													CGG Ala	·	432
10				AGG Arg										•	480
		GGA Gly		AAT Asn	-	GCC Ala								,	528
15				TTC Phe 180										:	576
20				CCC Pro										(624
				CAG Gln										(672
25 .				GAT Asp										7	720
30				CTA Leu										7	768
35				GAA Glu 260										8	316
				AGC Ser											364
	Asn			CCC Pro											912
45				CTT Leu				TAG	-					g	36
50				N FO											
	(i) SEC	UEN(CE CH	IARA	CTER	ISTIC	S:							

- (A) LENGTH: 918 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

50

55

(A) NAME/KEY: Coding Sequence (B) LOCATION: 1...915

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

. 10												GAA Glu					. 48	}
. 15												AGA Arg				AGG Arg	96	
•												GGA Gly				GAT Asp	144	1
20 .	ATA Ile											GCT Ala 60					192	
25	Pro 65	Lys	Ser	Ser	Gly	Pro 70	Tyr	Gly	Val	Leu	Val 75	TAT Tyr	Leu	His	Gly	Gly 80	240	•
												CCA Pro					288	•
30												GTG Val					336	•
35												GAT Asp					384	•
40												GAT Asp 140					432	
	GTT Val 145											TTG Leu					480	•
45												TAT Tyr					528	

,									GTT Val 185							TAC Tyr	576
5									GAG Glu								624
10									TTG Leu								672
15									TTA Leu								720
									CAA Gln								768
20									ACT Thr 265								816
25									CCG Pro								864
<i>30</i>									TTA Leu								912
30	ATT Ile	TAA															918
35	(2) INF		ATION JENCI					S:									
40		(B) 1	LENG [*] FYPE: FOPO	amin	o acid	i	cids										
	(ii)	MOLI	ECUL	E TYF	PE: pr	otein											
45	(v)	FRAC	SMEN	IT TYI	PE: in	ternal											
	(xi)	SEQ	UENC	E DE	SCRI	PTIO	N: SE	Q ID	NO:33	l :							
50	Met 1	Ser	Leu	Asn	Lys 5	His	Ser	Trp	Met	Asp 10	Met	Ile	Ile	Phe	Ile 15	Leu	
	Ser	Phe	Ser	Phe 20	Pro	Leu	Thr	Met	Ile. 25		Leu	Ala	Ile	Ser 30		Ser	
55	Ser	Trp	Phe 35	Asn	Ile	Trp	Asn	Asn 40	Ala	Leu	Ser	Asp	Leu 45	Gly	His	Ala	

	Val	Lys 50	Ser	Ser	Val	Ala	Pro 55		Phe	Asn	Leu	Gly 60	Leu	Ala	Ile	Gly
5	Gly 65	Ile	Leu	Ile	Val	Ile 70	Val	Gly	Leu	Arg	Asn 75	Leu	Tyr	Ser	Trp	Ser 80
	Arg	Val	Lys	Gly	Ser 85	Leu	Ile	Ile	Ser	Met 90	Gly	Val	Phe	Leu	Asn 95	Leu
	Ile	Gly	Val	Phe 100	Asp	Glu	Val	Tyr	Gly 105	Trp	Ile	His	Phe	Leu 110	Val	Ser
10 .	Val	Leu	Phe 115	Phe	Leu	Ser	Ile	Ile 120	Ala	Tyr	Phe	Ile	Ala 125	Ile	Ser	Ile
	Leu	Asp 130	Lys	Ser	Trp	Ile	Ala 135	Val	Leu	Leu	Ile	Ile 140	Gly	His	Ile	Ala
	Met 145	Trp	Tyr	Leu	His	Phe 150	Ala	Ser	Glu	Ile	Pro 155	Arg	Gly	Ala	Ala	Ile 160
15	Pro	Glu	Leu	Leu	Ala 165	Val	Phe	Ser	Phe	Leu 170	Pro	Phe	Tyr	Ile	Arg 175	Asp
	Tyr	Phe	Lys	Ser 180	Tyr	Thr	Lys	Arg								

20

25

30

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 346 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

35

40

50

Met 1	Lys	Leu	Leu	Glu 5	Pro	Thr	Asn	Thr	Ser 10	Tyr	Thr	Leu	Leu	Gln 15	Asp
Leu	Ala	Leu	His 20	Phe	Ala	Phe	Tyr	Trp 25	Phe	Leu	Ala	Val	Tyr 30	Thr	Trp
Leu	Pro	Gly 35	Val	Leu	Val	Arg	Gly 40		Ala	Val	Asp	Thr 45	Gly	Val	Ala
Arg	Va1 50	Pro	Gly	Leu	Gly	Arg 55	Arg	Gly	Lys	Arg	Leu 60	Leu	Leu	Ala	Ala
Val 65	Ala	Val	Leu	Ala	Leu 70	Val	Val	Ser	Val	Val 75	Val	Pro	Ala	Tyr	Val 80
Ala	Tyr	Ser	Ser	Leu 85	His	Pro	Glu	Ser	Cys 90	Arg	Pro	Val	Ala	Pro 95	Glu
Gly	Leu	Thr	Tyr 100	Lys	Glu	Phe	Ser	Val 105	Thr	Ala	Glu	Asp	Gly 110	Leu	Val
Val	Arg	Gly 115	Trp	Val	Leu	Gly	Pro 120	Gly	Ala	Gly	Gly	Asn 125	Pro	Val	Phe
Val	Leu 130	Met	His	Gly	Tyr	Thr 135	Gly	Cys	Arg	Ser	Ala 140	Pro	Tyr	Met	Ala
Val 145	Leu	Ala	Arg	Glu	Leu 150	Val	Glu	Trp	Gly	Tyr 155	Pro	Val	Val	Val	Phe 160
Asp	Phe	Arg	Gly	His 165	Gly	Glu	Ser	Gly	Gly 170	Ser	Thr	Thr	Ile	Gly 175	Pro

•	Arg	Glu	Val	Leu 180	Asp	Ala	Arg	Ala	Val 185	Val	Gly	Tyr	Val	Ser 190		Arg
	Phe	Pro	Gly 195	Arg	Arg	Ile	Ile	Leu 200	Val	Gly	Phe	Ser	Met 205	Gly	Gly	Ala
	Val	Ala 210	Ile	Val	Glu	Gly	Ala 215	Gly	Asp	Pro	Arg	Val 220	Tyr	Ala	Val	Ala
	Ala 225	Asp	Ser	Pro	Tyr	Tyr 230	Arg	Leu	Arg	Asp	Val 235	Ile	Pro	Arg	Trp	Leu 240
	Glu	Tyr	Lys	Thr	Pro 245	Leu	Pro	Gly	Trp	Val 250	Gly	Val	Leu	Ala	Gly 255	Phe
	Tyr	Gly	Arg	Leu 260	Met	Ala	Gly	Val	Asp 265	Leu	Gly	Phe	Gly	Pro 270	Ala	Gly
	Val	Glu	Arg 275	Val	Asp	Lys	Pro	Leu 280	Leu	Val	Val	Tyr	Gly 285	Pro	Arg	Asp
	Pro	Leu 290	Val	Thr	Arg	Asp	Glu 295	Ala	Arg	Ser	Leu	Ala 300	Ser	Arg	Ser	Pro
• ,	Cys 305	Gly	Arg	Leu	Val	Glu 310	Val	Pro	Gly	Ala	Gly 315	His	Val	Glu	Ala	Val 320
	Asp	Val	Leu	Gly	Pro 325	Gly	Arg	Tyr	Ala	Asp 330	Met	Leu	Ile	Glu	Leu 335	Ala
	His	Glu	Glu	Cys 340	Pro	Pro	Gly	Ala	Gly 345	Gly						

(2) INFORMATION FOR SEQ ID NO:35:

15

20

25

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 262 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

40	Met 1	Pro	Tyr	Val	Arg 5	Asn	Gly	Gly	Val	Asn 10	Ile	Tyr	Tyr	Glu	Leu 15	Val
	Asp	Gly	Pro	Glu 20	Pro	Pro	Ile	Val	Phe 25	Val	His	Gly	Trp	Thr 30	Ala	Asn
	Met	Asn	Phe 35	Trp	Lys	Glu	Gln	Arg 40	Arg	Tyr	Phe	Ala	Gly 45	Arg	Asn	Met
45	Met	Leu 50	Phe	Val	Asp	Asn	Arg 55	Gly	His	Gly	Arg	Ser 60	Asp	Lys	Pro	Leu
	Gly 65	Tyr	Asp	Phe	Tyr	Arg 70	Phe	Glu	Asn	Phe	Ile 75	Ser	Asp	Leu	Asp	Ala 80
50	Val	Val	Arg	Glu	Thr 85	Gly	Val	Glu	Lys	Phe 90	Val	Leu	Val	Gly	His 95	Ser
30	Phe	Gly	Thr	Met 100	Ile	Ser	Met	Lys	Tyr 105	Cys	Ser	Glu	Tyr	Arg 110	Asn	Arg
	Val	Leu	Ala 115	Leu	Ile	Leu	Ile	Gly 120	Gly	Gly	Ser	Arg	Ile 125	Lys	Leu	Leu
55	His	Arg 130	Ile	Gly	Tyr	Pro	Leu 135	Ala	Lys	Ile	Leu	Ala 140	Ser	Ile	Ala	Tyr

	Lys 145	Lys	Ser	Ser	Arg	Leu 150	Val	Ala	Asp	Leu	Ser 155	Phe	Gly	Lys	Asn	Ala 160
	Gly	Glu	Leu	Lys	Glu 165	Trp	Gly	Trp	Lys	Gln 170	Ala	Met	Asp	Tyr	Thr 175	Pro
	Ser	Tyr	Val	Ala 180	Met	Tyr	Thr	Tyr	Arg 185	Thr	Leu	Thr	Lys	Val 190	Asn	Leu
	Glu	Asn	Ile 195	Leu	Glu	Lys	Ile	Asp 200	Cys	Pro	Thr	Leu	Ile 205	Ile	Val	Gly
	Glu	Glu 210	Asp	Ala	Leu	Leu	Pro 215	Val	Ser	Lys	Ser	Val 220	Glu	Leu	Ser	Arg
•	Arg 225	Ile	Glu	Asn	Ser	Lys 230	Leu	Val	Ile	Ile	Pro 235	Asn	Ser	Gly	His	Cys 240
	Val	Met	Leu	Glu	Ser 245	Pro	Ser	Glu	Val	Asn 250	Arg	Ala	Met	Asp	Glu 255	Phe
	Ile	Ser	Ser	Ala 260	Gln	Phe										

(2) INFORMATION FOR SEQ ID NO:36:

10

15

20

25

30

35

45

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 251 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Leu 1	Arg	Leu	Arg	Lys 5	Phe	Glu	Glu [.]	Ile	Asn 10	Leu	Val	Leu	Ser	Gly 15	Gly
Ala	Ala	Lys	Gly 20	Ile	Ala	His	Ile	Gly 25	Val	Leu	Lys	Ala	Ile 30	Asn	Glu
Leu	Gly	Ile 35	Arg	Val	Arg	Ala	Leu 40	Ser	Gly	Val	Ser	Ala 45	Gly	Ala	Ile
Val	Ser 50	Val	Phe	Tyr	Ala	Ser 55	Gly	Tyr	Ser	Pro	Glu 60	Gly	Met	Phe	Ser
Leu 65	Leu	Lys	Arg	Val	Asn 70	Trp	Leu	Lys	Leu	Phe 75	Lys	Phe	Lys	Pro	Pro 80
Leu	Lys	Gly	Leu	Ile 85	Gly	Trp	Glu	Lys	Ala 90	Ile	Arg	Phe	Leu	Glu 95	Glu
Val	Leu	Pro	Tyr 100	Arg	Arg	Ile	Glu	Lys 105	Leu	Glu	Ile	Pro	Thr 110	Tyr	Ile
Cys	Ala	Thr 115	Asp	Leu	Tyr	Ser	Gly 120	Arg	Ala	Leu	Tyr	Leu 125	Ser	Glu	Gly
Ser	Leu 130	Ile	Pro	Ala	Leu	Leu 135	Gly	Ser	Cys	Ala	Ile 140	Pro	Gly	Ile	Phe
Glu 145	Pro	Val	Glu	Tyr	Lys 150	Asn	Tyr	Leu	Leu	Val 155	Asp	Gly	Gly	Ile	Val 160
Asn	Asn	Leu	Pro	Val 165	Glu	Pro	Phe	Gln	Glu 170	Ser	Gly	Ile	Pro	Thr 175	Val
Cys	Val	Asp	Val 180	Leu	Pro	Ile	Glu	Pro 185	Glu	Lys	Asp	Ile	Lys 190	Asn	Ile
Leu	His	Ile 195	Leu	Leu	Arg	Ser	Phe 200	Phe	Leu	Ala	Val	Arg 205	Ser	Asn	Ser

Glu Lys Arg Lys Glu Phe Cys Asp Leu Val Ile Val Pro Glu Leu Glu 210 215 220

Glu Phe Thr Pro Leu Asp Val Arg Lys Ala Asp Gln Ile Met Glu Arg 225 230 230 240

Gly Tyr Ile Lys Ala Leu Glu Val Leu Ser Glu 250

(2) INFORMATION FOR SEQ ID NO:37:

10

15

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 297 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

25	Met 1	Phe	Asn	Ile	Asn 5	Val	Phe	Val	Asn	Ile 10	Ser	Trp	Leu	Tyr	Phe 15	Ser
	Gly	Ile	Val	Met 20	Lys	Thr	Val	Glu	Glu 25	Tyr	Ala	Leu	Leu	Glu 30	Thr	Gly
	Val	Arg	Val 35	Phe	Tyr	Arg	Cys	Val 40	Ile	Pro	Glu	Lys	Ala 45	Phe	Asn	Thr
30	Leu	Ile 50	Ile	Gly	Ser	His	Gly 55	Leu	Gly	Ala	His	Ser 60	Gly	Ile	Tyr	Ile
	Ser 65	Val	Ala	Glu	Glu	Phe 70	Ala	Arg	His	Gly	Phe 75	Gly	Phe	Суѕ	Met	His 80
	Asp	Gln	Arg	Gly	His 85	Gly	Arg	Thr	Ala	Ser 90	Asp	Arg	Glu	Arg	Gly 95	Tyr
35	Val	Glu	Gly	Phe 100	His	Asn	Phe	Ile	Glu 105	Asp	Met	Lys	Ala	Phe 110	Ser	Asp
	Tyr	Ala	Lys 115	Trp	Arg	Val	Gly	Gly 120	Asp	Glu	Ile	Ile	Leu 125	Leu	Gly	His
40	Ser	Met 130	Gly	Gly	Leu	Ile	Ala 135	Leu	Leu	Thr	Val	Ala 140	Thr	Tyr	Lys	Glu
	Ile 145	Ala	Lys	Gly	Val	Ile 150	Ala	Leu	Ala	Pro	Ala 155	Leu	Gln	Ile	Pro	Leu 160
	Thr	Pro	Ala	Arg	Arg 165	Leu	Val	Leu	Ser	Leu 170	Ala	Ser	Arg	Leu	Ala 175	Pro
45	His	Ser	Lys	Ile 180	Thr	Leu	Gln	Arg	Arg 185	Leu	Pro	Gln	Lys	Pro 190	Glu	Gly
	Phe	Gln	Arg 195	Ala	Lys	Asp	Ile	Glu 200	Tyr	Ser	Leu	Ser	Glu 205	Ile	Ser	Val
	Lys	Leu 210	Val	Asp	Glu	Met	Ile 215	Lys	Ala	Ser	Ser	Met 220	Phe	Trp	Thr	Ile
50	Ala 225	Gly	Glu	Ile	Asn	Thr 230	Pro	Val	Leu	Leu	Ile 235	His	Gly	Glu	Lys	Asp 240
	Asn	Val	Ile	Pro	Pro 245	Glu	Ala	Ser	Lys	Lys 250	Ala	Tyr	Gln	Leu	Ile 255	Pro
	Ser	Phe	Pro	Lys 260	Glu	Leu	Lys	Ile	Tyr 265	Pro	Asp	Leu	Gly	His 270	Asn	Leu
55																

Phe Phe Glu Pro Gly Ala Val Lys Ile Val Thr Asp Ile Val Glu Trp Val Lys Asn Leu Pro Arg Glu Asn Pro

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 262 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Met Glu Val Tyr Lys Ala Lys Phe Gly Glu Ala Lys Leu Gly Trp Val Val Leu Val His Gly Leu Gly Glu His Ser Gly Arg Tyr Gly Arg Leu Ile Lys Glu Leu Asn Tyr Ala Gly Phe Gly Val Tyr Thr Phe Asp Trp Pro Gly His Gly Lys Ser Pro Gly Lys Arg Gly His Thr Ser Val Glu Glu Ala Met Glu Ile Ile Asp Ser Ile Ile Glu Glu Ile Arg Glu Lys Pro Phe Leu Phe Gly His Ser Leu Gly Gly Leu Thr Val Ile Arg Tyr Ala Glu Thr Arg Pro Asp Lys Ile Arg Gly Leu Ile Ala Ser Ser Pro Ala Leu Ala Lys Ser Pro Glu Thr Pro Gly Phe Met Val Ala Leu Ala Lys Phe Leu Gly Lys Ile Ala Pro Gly Val Val Leu Ser Asn Gly Ile Lys Pro Glu Leu Leu Ser Arg Asn Arg Asp Ala Val Arg Arg Tyr Val Glu Asp Pro Leu Val His Asp Arg Ile Ser Ala Lys Leu Gly Arg Ser Ile Phe Val Asn Met Glu Leu Ala His Arg Glu Ala Asp Lys Ile Lys Val Pro Ile Leu Leu Ile Gly Thr Gly Asp Val Ile Thr Pro Pro Glu Gly Ser Arg Arg Leu Phe Glu Glu Leu Ala Val Glu Asn Lys Thr Leu Arg Glu Phe Glu Gly Ala Tyr His Glu Ile Phe Glu Asp Pro Glu Trp Ala Glu Glu Phe His Glu Thr Ile Val Lys Trp Leu Val Glu Lys Ser Tyr Ser Ser Ala Gln

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 249 amino acids
- (B) TYPE: amino acid (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

10 Leu Ile Gly Asn Leu Lys Leu Lys Arg Phe Glu Glu Val Asn Leu Val 1 10 Leu Ser Gly Gly Ala Ala Lys Gly Ile Ala His Ile Gly Val Leu Lys 20 25 15 Ala Leu Glu Glu Leu Gly Ile Lys Val Lys Arg Leu Ser Gly Val Ser 40 Ala Gly Ala Ile Val Ser Val Phe Tyr Ala Ser Gly Tyr Thr Pro Asp 55 50 60 Glu Met Leu Lys Leu Lys Glu Val Asn Trp Leu Lys Leu Phe Lys 20 70 Phe Lys Thr Pro Lys Met Gly Leu Met Gly Trp Glu Lys Ala Ala Glu 85 90 Phe Leu Glu Lys Glu Leu Gly Val Lys Arg Leu Glu Asp Leu Asn Ile 100 105 110 Pro Thr Tyr Leu Cys Ser Ala Asp Leu Tyr Thr Gly Lys Ala Leu Tyr 25 115 120 125 Phe Gly Arg Gly Asp Leu Ile Pro Val Leu Gly Ser Cys Ser Ile 130 135 140 Pro Gly Ile Phe Glu Pro Val Glu Tyr Glu Asn Phe Leu Leu Val Asp 150 155 Gly Gly Ile Val Asn Asn Leu Pro Val Glu Pro Leu Glu Lys Phe Lys 165 170 175 Glu Pro Ile Ile Gly Val Asp Val Leu Pro Ile Thr Gln Glu Arg Lys 180 185 Ile Lys Asn Ile Leu His Ile Leu Ile Arg Ser Phe Phe Leu Ala Val 35 195 200 205 Arg Ser Asn Ser Glu Lys Arg Lys Glu Phe Cys Asn Val Val Ile Glu 210 215 220 Pro Pro Leu Glu Glu Phe Ser Pro Leu Asp Val Asn Lys Ala Asp Glu 230 235 Ile Phe Cys Gly Asp Met Arg Ala Leu

- (2) INFORMATION FOR SEQ ID NO:40:
- 45 (i) SEQUENCE CHARACTERISTICS:

40

50

(A) LENGTH: 338 amino acids

- (B) TYPE: amino acid
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- 55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

	Met 1	Pro	Ala	Asn	Asp 5	Ser	Pro	Thr	Ile	Asp 10	Phe	Asn	Pro	Arg	Gly 15	Ile
5	Leu	Arg	Asn	Ala 20	His	Ala	Gln	Val	Ile 25	Leu	Ala	Thr	Ser	Gly 30	Leu	Arg
	Lys	Ala	Phe 35	Leu	Lys	Arg	Thr	His 40	Lys	Ser	Tyr	Leu	Ser 45	Thr	Ala	Gln
	Trp	Leu 50	Glu	Leu	Asp	Ala	Gly 55	Asn	Gly	Val	Thr	Leu 60	Ala	Gly	Glu	Leu
10	Asn 65	Thr	Ala	Pro	Ala	Thr 70	Ala	Ser	Ser	Ser	His 75	Pro	Ala	His	Lys	Asn 80
				-	85	Leu		_	_	90	_				95	
	-			100		Gly			105		_		_	110	_	
15		_	115			Arg	_	120	_	, -		_	125			_
•	_	130				Ser	135		_			140	_			_
20	145					Thr 150	_	_	_	=	155	_			-	160
			-	•	165	Phe			-	170			-		175	
				180		Ala			185					190		
25			195			Met		200		-	-		205	-	_	_
	-	210				Lys	215		_			220			•	
	225				-	Lys 230				_	235					240
30		•			245	Asn Phe	-			250	-	-		-	255	
				260	-	Pro	-		265				_	270	_	
35			275		•	Asp		280					285	-	•	
		290				Gln	295		-			300				
	305			•		310 Ala		-			315		•			320
40	Ala		-,5	-10	325					330			_,,		335	 1
•		- 3 -														

(2) INFORMATION FOR SEQ ID NO:41:

45

50

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 311 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

	Met I	Leu As _i) Met	Pro 5	Ile	Asp	Pro	Val	Tyr 10	Tyr	Gln	Leu	Ala	Glu 15	Tyr
5		Asp Se	20		•		_	25					зŏ		•
	_	Slu Ala 3	5		-		40			_		45			
		lis Glu 50				55			-	-	60		-	_	-
10	65	Sly As			70					75			-		80
		eu Vai		85					90					95	
		lis Ası	100		-	•	_	105					110		
15		al Ser 11	,		•	_	120					125			
	1	/al Ty: 130				135				-	140				
20	145	Slu Lei	-		150			_		155		_	-	•	160
		Sly Gly		165					170				_	175	
	-	Slu Asp	180	-	•			185			-		190		
25		al Ala 19	i				200				-	205	-		-
	2	eu Ası 10		-		215		=			220		-	• • •	•
	225	ilu Glı	_	_	230					235					240
30		ilu Ası		245					250				-	255	
		rg Asp	260					265				_	270		_
35	,	31u Ala 279	•				280					285	_		
	2	'yr Ty: 90				295	Ala	Ala	Arg	Asp	Ala 300	Ile	Asn	Gln	Ile
	Ala A 305	la Let	Leu	Val	Phe 310	Asp									

40

45

50

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 305 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

	Met 1	Pro	Leu	Asp	Pro 5	Arg	Ile	Lys	Lys	Leu 10	Leu	Glu	Ser	Ala	Leu 15	Thr
5	Ile	Pro	Ile	Gly 20	Lys	Ala	Pro	Val	Glu 25	Glu	Val	Arg	Lys	Ile 30	Phe	Arg
	Gln	Leu	Ala 35	Ser	Ala	Ala	Pro	Lys 40	Val	Glu	Val	Gly	Lys 45	Val	Glu	Asp
	Ile	Lys 50	Ile	Pro	Gly	Ser	Glu 55	Thr	Val	Ile	Asn	Ala 60	Arg	Val	Tyr	Phe
10	Pro 65	Lys	Ser	Ser	Gly	Pro 70	Tyr	Gly	Val	Leu	Val 75	Tyr	Leu	His	Gly	Gly 80
	Gly	Phe	Val	Ile	Gly 85	Asp	Val	Glu	Ser	Tyr 90	Asp	Pro	Leu	Cys	Arg 95	Ala
45	Ile	Thr	Asn	Ala 100	Cys	Asn	Cys	Val	Val 105	Val	Ser	Val	Asp	Tyr 110	Arg	Leu
15	Ala	Pro	Glu 115	Tyr	Lys	Phe	Pro	Ser 120	Ala	Val	Ile	Asp	Ser 125	Phe	Asp	Ala
	Thr	Asn 130	Trp	Val	Tyr	Asn	Asn 135	Leu	Asp	Lys	Phe	Asp 140	Gly	Lys	Met	Gly
20	Val 145		Ile	Ala	Gly	Asp 150	Ser	Ala	Gly	Gly	Asn 155	Leu	Ala	Ala		Val 160
			Leu		165					170		_			175	-
	Tyr	Pro	Ala	Val 180	Ser	Leu	Asp	Asn	Val 185	Ser	Arg	Ser	Met	Ile 190	Glu	Tyr
25		-	Gly 195					200					205		•	
		210	Leu				215					220	_			
	225		Ala		_	230		-			235					240
30			Tyr	_	245		-	_		250			•		255	-
			Gln	260					265			_		270		
35			Gly 275					280					285	_	_	_
		11e 290	Gly	Leu	Ile	Gly	Ser 295	Val	Leu	Arg	Arg	Val 300	Phe	Tyr	Asp	Lys
	11e 305															
40										•						

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 605 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (ix) FEATURE:

45

50

- (A) NAME/KEY: Coding Sequence
- (B) LOCATION: 1...603
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

		AAG Lys						GTT Val									48
5 .								TCG Ser								AAA Lys	96
10								ACT Thr 40									144
15								ATT Ile									192
20								GTT Val									240
20								GAA Glu									288
25								GAA Glu									336
30								GAG Glu 120									384
								TAT Tyr									432
35								AAA Lys									480
40								AAG Lys									528
45	CAT	TTA	AGC	CAT	ACA	AGT	ATG	ATT	TAC	TCT	CGC	CAA	GTC	GTT	AAT	TAT	576
	His	Leu	Ser	His 180	Thr	Ser	Met	Ile	Tyr 185	Ser	Arg	Gln	Val	Val 190	Asn	Tyr	·
50								GAC Asp 200		TA	-						605

(2) INFORMATION FOR SEQ ID NO:44:

55

(i) SEQUENCE CHARACTERISTICS:

_		(B) T (C) S	ENGT YPE: I TRAN OPOL	nucleio DEDN	acid IESS:	single									
•	(ii) l	MOLE	CULE	TYPE	E: Ger	omic	DNA								
	(ix)	FEAT	URE:												
0			AME/I				uence								
	(xi)	SEQL	JENCI	E DES	CRIP	TION:	SEQ	ID NO	D:44:						
5			AAA Lys										-		 48
	1				5					10				15	
20			GCT Ala												96
25			AAT Asn 35												 144
			GGT Gly									Lys			192
			ATC Ile												240
35			GGT Gly												288
10			CTT Leu												 336
			ATC Ile 115										_		 384
4 5									ſ						

5		TCG Ser 130	Asp														432
		CCC															480
10	-	GGT Gly															528
15		GTT Val															576
20		CCG Pro															624
		TCA Ser 210															672
25		GCA Ala									•						720
30		ACA Thr															768
. 35		TTA Leu		TG													779
	(2) INF	ORMA	TION	FOR S	SEQ II	ONO:	45:										
	(i) S	SEQUE	ENCE	CHAF	RACTI	ERIST	ICS:										
40		(B) T (C) S	ENGT YPE: r TRAN OPOL	nucleio DEDN	acid IESS:	single		•							•		
45	(ii)	MOLE	CULE	TYPE	: Gen	omic	DNA										
	(ix)	FEAT	URE:														
50			AME/I		_		ience										
	(xi)	SEQL	JENCE	DES	CRIP	TION:	SEQ	ID NC	:45:								
55	ATG	ccc	СТА	CAT	CCA	AAG	GTA	AAG	AAA	TTA	СТТ	TCC	CAG	CTA (CCT	ccc	48

	Met 1	Pro	Leu	His	Pro 5	Lys	Val	Lys	Lys ·	Leu 10	Leu	Ser	Gln	Leu	Pro 15	Pro	
5																TTA Leu	96
10		TTC Phe														GAG Glu	144
15		CCC Pro 50														TCA Ser	192
		AAG Lys															240
20		TTC Phe															288
25		GAA Glu															336
30		CAC His															384
		ATC Ile 130															432
35		GTG Val															480
40		CTG Leu					_										528
45		TAC Tyr															576
		GGC Gly															624
50		ATG Met 210															672

FP 0 880 590 R1

٠.	Pro	Ala				Leu	His			Pro	Ser			Thr	. 720
5	225					230				 235				240	
		GAG Glu				Leu									768
10		AAC Asn			Gly								Gly		816
15		CAC His													864
20		Ile 290											٠		905
	(2) INF	ORMA'	TION I	FOR S	SEQ II	D NO:	46:								
25	(i) S	SEQUE	ENCE	CHAF	RACTI	ERIST	FICS:								
		(B) T (C) S	ENGTI YPE: r TRAN	nucleid DEDN	acid IESS:	single									
30	(ii)	MOLE	OPOL CULE				DNA								
	(ix)	FEAT	URE:												
35			AME/H DCATI				uence								•
	(xi)	SEQU	IENCE	DES	CRIP	TION:	SEQ	ID NO	D:46:						
40		GCC Ala													48
45		AGG Arg													96
50		GAA Glu													144
		CAT His 50													192
55															

5						AAG Lys 70											240
						GTC Val											288
10						TGG Trp											336
15	Phe	Val	Met 115	Asp	Thr	AGG Arg GAG	Gly	Gln 120	Gly	Ser	Gly	Trp	Met 125	Lys	Gly	Asp	384
	Thr					Glu											432
20						CTG Leu 150										GTC Val 160	480
25						AGG Arg											528 .
30						GTG Val											576
						AGT Ser											624
35						CTG Leu											672
40						GTG Val 230											720
45						GTT Val					Ser						768
						GCA Ala											816
50						CCT Pro											864

5	Tyr A	CC GGT la Gly 90											912
		GA GGT ly Gly											960
10		TT GAG he Glu			TAG							·	978
15	(2) INFORI	MATION	FOR S	SEQ II	D NO:	47:							
	(i) SEC	QUENCE	CHAF	RACTI	ERIST	TICS:					2		
20	(B (C) LENG1) TYPE:) STRAN) TOPOI	nucleic NDEDN	acid IESS:	single								٠
	(ii) MC	LECULE	TYPE	: Ger	omic	DNA							
25	(ix) FE	ATURE:											
	•) NAME/) LOCAT				uence							
30	(xi) SE	QUENC	E DES	CRIP	TION:	SEQ	ID NO):47:					
35	ATG CO Met A												48
	ATG GO												96
40	CAG GC Gln Gl												144
45	CGC AT												192
50	CTC GC Leu Al 65												240
	CAC GG His Gl												288
55	CTG GC Leu Al												336

5												GGA Gly	384
10		GCC Ala 130								Ile			432
		GCC Ala											480
15		AAG Lys											528
20		CGC Arg											576
25		AAG Lys									 		624
23		CTC Leu 210											672
30		CAG Gln											720
35		TTC Phe											768
		CTG Leu											816
40		CTC Leu											864
45		TTC Phe 290							TAG				897
50	(2) INFO	ORMA	TION	FOR :	S EQ I	D NO:	48:						
	(i) S	EQUI	ENCE	CHAI	RACT	ERIST	TICS:						

(ii) MOLECULE TYPE: Genomic DNA

(A) LENGTH: 914 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Coding Sequence (B) LOCATION: 1...912

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

10		AAA TT												48
15	GAA CCI Glu Pro	GTG AT Val Me 2	t Lys :											96
	GCA CGG Ala Arg					Gln								144
20	GCA ACA Ala Thr 50	Arg Pr			llle									192
25	ATG ACG Met Thr 65													240
30	TTC CGT Phe Arg		*											288
	GCC CGT Ala Arg		r Ser (336 .
35	GTG TAT Val Tyr					Val							_	384
40	GAC GGG Asp Gly 130	Val Cy			ı Ala									432
·	TCG GTC Ser Val 145		r Arg 1											480
45	CTT GAT Leu Asp												-	528
50	CTT GGC	GGC GA	CCT (GAA CG	r atc	GGC	GTT	GGC	GGC	GAT	AGC	GCA	GGG	576

	Leu	Gly	Gly	Asp 180		Glu	Arg	Ile	Gly 185	Val	Gly	Gly	Asp	Ser 190	Gly		
.		AAT Asn														624	į
10		ACA Thr 210	Pro													672	!
15		ATC Ile														720)
		TCC Ser														768	1
20		ATC Ile														816	;
25		CAA Gln														864	
30		CGG Arg 290														912	
	TA															914	
35	(2) INF	ORMA SEQUI															
40		(B) T (C) S	ENGT YPE: TRAN	nuclei IDEDN	c acid NESS:	· single											
	(ii)	MOLE	CULE	TYP	E: Ger	nomic	DNA										
45	(ix)	FEAT	URE:														
			IAME/I				uence										
50	(xi)	SEQL	JENC	E DES	CRIP	TION:	SEQ	ID NO):49:								
55		AGC Ser														48	
		CCA Pro														96	

*					•								•				
				-20					25					30			
5	TTT Phe	GAA Glu	TTA Leu 35	Lys	GCG Ala	AGG Arg	TTT	TTG Leu 40	TTT Phe	CCG Pro	GCG Ala	CCA Pro	CGT Arg 45	AAA Lys	ACA Thr	AGG Arg	144
10					GTA Val												192
					AAA Lys												240
15	CAT His	GG GG	GGA Gly	GGG Gly	TAT Tyr 85	GTT Val	TTT Phe	GGG	TCA Ser	CCA Pro 90	AAA Lys	ACG Thṛ	CAC His	CGT Arg	GCA Ala 95	ATG Met	288
20					TCG Ser												336
25					CCA Pro												38.4
	GTT Val	TTA Leu 130	TCG Ser	TAT Tyr	AAA Lys	TGT Cys	TTA Leu 135	CTA Leu	GAG Glu	CGA Arg	GCA Ala	ATC Ile 140	GAG Glu	CCC Pro	CAA Gln	AAT Asn	432
30					GGG Gly												480
35					AAG Lys 165												528
40					TTG Leu												576
					ACC Thr												624
45					TAT Tyr												672
50		Pro			GCG Ala												720
	GCA	AGT	GAC	agt	GAA	ATC	CTG	TTG	GAT	GAT	TGC	CTG	CGG	ATG	GCG	GAT	768

	Ala	Ser	Asp	Ser	G1u 245		Leu	Leu	Asp	250	-	Leu	Arg	Met	25!	a Asp		
5					Gln					Thr					Gli	A AAC u Asn	816	
10				Val					Gln					Glu		A GAT	864	
15			Leu					Ala					Leu			A GGT c Gly	912	
			GAA Glu	AGC Ser												•	926	
20	(2) INFC)RMA	TION	FOR S	SEQ II	D NO:	50·											
	(1) S	EQUI	ENCE	CHA	RACTI	ERIST	ICS:											
		(B) T (C) S	YPE: i TRAN	nuclei IDEDN	3 base c acid NESS: linear	single		•										
30	(ii)	MOLE	CULE	TYPE	E: Ger	nomic	DNA											
	(ix)	FEAT	URE:															
35					Coding		Jence											
	(xi)	SEQL	JENCI	E DES	SCRIP	TION:	SEQ	ID NO	D:50:									
40					AAT Asn 5												48	
45					TTT Phe												96	
50					AAA Lys												144	
50					CGG Arg												192	
55					CAG Gln												240	

	•															
5								ACA Thr								288
								CAG Gln								336
10								AAT Asn 120			Met					384
15			GCT					ATG			GCG	CAA			GGG Gly	432
	Ile	ACG				Суз	ACC	GTG Val		Ala	GAT				Trp	480
20								GAA Glu								528
a -					165			CAG	170					175		576
25								Gln								3.0
30								GGA Gly 200								624
								GCA Ala								672
<i>3</i> 5								GAG Glu					TA			713
40	(2) INFO	RMAT	'ION F	FOR S	EQ IC	NO:	51:									
		EQUE (A) LE				RIST	ICS:									
45	í	(B) TY	PE: n	iucleic DEDN	acid ESS:	single	!									
50	(ii) M	OLEC	CULE	TYPE	: Gen	omic l	DNA									
	(ix) F	EATL	JRE:													
		(A) NA (B) LC				Sequ	ence						-			

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

5 .													,			CCA Pro	48
10																CTC Leu	96
15																GAG Glu	144
. 											GTC Val						192
20	AGG Arg 65										GTT Val 75						240
25											GGA Gly						288
											TCT Ser						336
30											GGC Gly						384
35											CCT Pro						432
40											TAC Tyr 155						480
											GCT Ala						528
4 5											GGC Gly						576
50											AAA Lys						624
	TGC	GAT	GTG	CCG	TTT	CTG	TGT	CAC	TTC	AGA	AGA	GCA	GTA	CAG	CTT	GTG	672

	Cys	Asp 210	Val	Pro	Phe	Leu	Cys 215	His	Phe	Arg	Arg	Ala 220	Val	Gln	Leu	Val	
5		ACG Thr															720
10	_	AAG Lys															768
		TTC Phe															816
15		GAC Asp															864
20		GCT Ala 290															912
25		GGA Gly															960
		TTT Phe				TAA											978
30	(2) INF(ORMA	TION	FOR S	SEQ II	D NO :	52:										
	(i) S	SEQUI	ENCE	CHA	RACTI	ERIST	ICS:										
35		(B) T (C) S	YPE: I	nuclei IDEDN	l base acid IESS: linear	single											
40	(ii) l	MOLE	CULE	TYPE	E: Ger	omic	DNA										
	(ix)	FEAT	URE:														
45					Codino 630		ience					-					
	(xi)	SEQU	JENCI	E DES	CRIP	TION:	SEQ	ID NO):52:								
50		AAG Lys															48
55		GCC Ala															96

5							TCC Ser 40						 144
10							TCC Ser						192
·							CGC Arg						240
15							CGC Arg						288
20							CTG Leu						336
25							CGC Arg 120						 384
25							TAT Tyr						432
30							TTG Leu						 480
.) 35							AAC Asn						528
							AGC Ser						576
40							CGC Arg 200						624
45	CAA Gln 210	TGAG	GGT	GAC (GTGC	CACCO	GA AC	GTGC	CACCI	G			661

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 201 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Met Lys Val Lys His Val Ile Val Leu His Gly Leu Tyr Met Ser Gly 10 Leu Val Met Arg Pro Leu Cys Ser Arg Leu Glu Glu Ser Gly Val Lys 20 25 Val Leu Asn Leu Thr Tyr Asn Thr Arg Asp Pro Asn Arg Asp Ala Ile 10 40 Phe Thr Gln Ile Asp Glu Phe Ile Ser Asn Glu Pro Ser Ala Leu Val 55 60 Cys His Ser Met Gly Gly Leu Val Ala Arg Ala Tyr Leu Glu Ala Asn 70 75 15 Ser Ala Pro Ser His His Val Glu Lys Val Ile Thr Leu Gly Thr Pro 85 90 His Thr Gly Ser His Ile Ala Glu Lys Met Gln Gln Lys Gly Phe Glu 100 105 110 Leu Leu Lys Asn Ser Val Glu Phe Leu Leu Ser Lys Asn Gly Asp 120 125 Trp Pro Phe Lys Ala Lys Leu Tyr Ser Ile Ala Gly Asp Leu Pro Ile 135 140 Gly Leu Met Pro Leu Ile Val Lys Gly Ser Arg Ser Asp Gly Thr Val 150 155 Leu Leu Asp Glu Thr Lys Leu Lys Gly Met Ala Glu His Lys Val Phe 170 165 His Leu Ser His Thr Ser Met Ile Tyr Ser Arg Gln Val Val Asn Tyr 180 185 Ile Leu Glu Arg Leu Asn Glu Asp Ile

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 259 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Met Ile Lys Asn Phe Asp Arg Glu Asn Ser Ser Leu Val Leu Ser Gly 1 5 10 15

Gly Gly Ala Leu Gly Ile Ala His Leu Gly Val Leu His Asp Leu Glu 25 30

Lys Gln Asn Ile Val Pro Asn Glu Ile Val Gly Thr Ser Met Gly Gly 35 40 45

Ile Ile Gly Ala Ser Met Ala Ile Gly Met Lys Glu Lys Glu Ile Leu 50 55 60

Glu Glu Ile Lys Asn Phe Ser Asn Val Phe Asn Trp Ile Lys Phe Ser

55

30

35

45

	65					70					75					80
P. 5	he S	er (Sly	Asn	Ser 85	Val	Val	Asp	Asn	Glu 90	Lys	Ile	Ala	Lys	Ile 95	Phe
A	sp T	hr I	Leu	Phe 100	Lys	Asp	Arg	Lys	Met 105	Thr	Asp	Thr	Val	Ile 110	Pro	Leu
L	ys L		[le [15	Ala	Thr	Asn	Leu	His 120	Asn	Gly	His	Lys	Lys 125	Val	Phe	Thr
10 A.		er <i>F</i> .30	Asp	Asp	Val	Leu	Ile 135	Lys	Asp	Ala	Ile	Leu 140	Ser	Thr	Met	Ala
	le P 45	ro G	Sly	Val		Glu 150	Glu	His	Ile	Ile	Asp 155	Gly	Glu	Thr	Tyr	Gly 160
A	sp G	ly F	?he	Leu	Cys 165	Glu	Asn	Leu	Gly	Val 170	Asn	Glu	Ala	Thr	Phe 175	Asn
15 A	sp V	al I		Ala 180	Val	Asp	Val	Met	Gly 185	Glu	Asn	Ser	Phe	Glu 190	Lys	Ala
Mo	et P		Asp 195	Asn	Phe	Phe	Lys	Thr 200	Ser	Asn	Val	Leu	Glu 205	Met	Phe	Glu
20		er M 10	let	Arg	Leu	Phe	Ile 215	Tyr	Asn	Gln	Thr	Gln 220	Thr	His	Ile	Lys
A:	sn A 25	la A	lsn	Lys	Asn	Ile 230	Tyr	Leu	Ile	Glu	Pro 235	Val	Thr	Lys	Glu	Tyr 240
L	ys T	hr F	Phe	Gln	Phe 245	His	Lys	His	Lys	Glu 250	Ile	Arg	Ala	Leu	Gly 255	Leu
G:	ly L	eu L	eu													

(2) INFORMATION FOR SEQ ID NO:55:

30

35

45

50

55

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 301 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Met Pro Leu His Pro Lys Val Lys Lys Leu Leu Ser Gln Leu Pro Pro 1 10 15 Gln Asp Phe Ser Arg Asn Val Gln Asp Leu Arg Lys Ala Trp Asp Leu 20 25 30 Pro Phe Ser Gly Arg Arg Glu Thr Leu Lys Arg Val Glu Asp Leu Glu 40 Ile Pro Thr Arg Asp Ala Arg Ile Arg Ala Arg Val Tyr Thr Pro Ser 55 Ser Lys Glu Asn Leu Pro Val Leu Val Tyr Tyr His Gly Gly Phe 70 75 Val Phe Gly Ser Val Asp Ser Tyr Asp Gly Leu Ala Ser Leu Ile Ala 90 85 Lys Glu Ser Gly Ile Ala Val Ile Ser Val Glu Tyr Arg Leu Ala Pro 105 Glu His Lys Phe Pro Thr Ala Val Asn Asp Ser Trp Asp Ala Leu Leu

	m 7	115	Glu Asr	Gly	G1 v	120 Lvs	Leu	Clv	Lan) en	125	Ser	Ara	T.O.I.
	-	30 Te Ara	GIU ASI	GIY	135	БуЗ	beu	GIY	Leu	140	1111	Jet	nry	Dea
5	145		Gly Asp	150					155					160
	Leu L	eu Asp	Arg Asp 165		Gly	Lys	Gly	Leu 170	Val	Ser	Tyr	Glņ	Val 175	Leu
10	Ile T	yr Pro	Ala Val	Asn	Met	Val	Asp 185	Asn	Ser	Pro	Ser	Val 190	Arg	Glu
	-	195	Gly Tyr			200					205			
•	2	10	Phe Ser		215					220				
15	Pro A 225	la Leu	Ala Asp	Leu 230	His	Asn	Leu	Pro	Pro 235	Ser	Leu	Val	Ile	Thr 240
	Ala G	lu Tyr	Asp Pro		Arg	Asp	Gln	Gly 250	Glu	Thr	Tyr	Ser	His 255	Ser
			Ala Gly 260/				265					270		
20	Ile H	is Gly 275	Phe Lei	Ser	Phe	Tyr 280	Glu	Trp	Ile	Thr	Ala 285	Gly	Lys	Leu
		le His 90	His Ile	Ala	Gly 295	Val	Leu	Arg	Ser	Val 300	Leu			

(2) INFORMATION FOR SEQ ID NO:56:

25

30

35

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 325 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:
- 40 Val Ala Phe Phe Asp Met Pro Leu Glu Glu Leu Lys Lys Tyr Arg Pro 10 Glu Arg Tyr Glu Glu Lys Asp Phe Asp Glu Phe Trp Arg Glu Thr Leu 25 20 Lys Glu Ser Glu Gly Phe Pro Leu Asp Pro Val Phe Glu Lys Val Asp 40 45 Phe His Leu Lys Thr Val Glu Thr Tyr Asp Val Thr Phe Ser Gly Tyr 55 Arg Gly Gln Arg Ile Lys Gly Trp Leu Leu Val Pro Lys Leu Ala Glu 75 Glu Lys Leu Pro Cys Val Val Gln Tyr Ile Gly Tyr Asn Gly Gly Arg 90 85 Gly Phe Pro His Asp Trp Leu Phe Trp Pro Ser Met Gly Tyr Ile Cys 105 110 100 Phe Val Met Asp Thr Arg Gly Gln Gly Ser Gly Trp Met Lys Gly Asp 125 120 Thr Pro Asp Tyr Pro Glu Gly Pro Val Asp Pro Gln Tyr Pro Gly Phe

		130					135				•	140				
_	Met 145	Thr	Arg	Gly	Ile	Leu 150	Asp	Pro	Gly	Thr	Tyr 155	Tyr	Tyr	Arg	Arg	Val 160
5	Phe	Val	Asp	Ala	Va,1 165	Arg	Ala	Val	Glu	Ala 170	Ala	Ile.	Ser	Phe	Pro 175	Arg
,	Val	Asp	Ser	Arg 180	Lys	Val	Val	Val	Ala 185	Gly	Gly	Ser	Gln	Gly 190	Gly	Gly
10	Ile	Pro	Leu 195	Ala	Val	Ser	Ala	Leu 200	Ser	Asn	Arg	Val	Lys 205	Ala	Leu	Leu
	Cys	Asp 210	Val	Pro	Phe	Leu	Cys 215	His	Phe	Arg	Arg	Ala 220	Val	Gln	Leu	Val
	Asp 225	Thr	His	Pro	Tyr	Val 230	Glu	Ile	Thr	Asn	Phe 235	Leu	Lys	Thr	His	Arg 240
15	Asp	Lys	Glu	Glu	Ile 245	Val	Phe	Arg	Thr	Leu 250	Ser	Tyr	Phe	Asp	Gly 255	Val
	Asn	Phe	Ala	Ala 260	Arg	Ala	Lys	Val	Pro 265	Ala	Leu	Phe	Ser	Val 270	Gly	Leu
	Met	Asp	Thr 275	Ile	Суѕ	Pro	Pro	Ser 280	Thr	Val	Phe	Ala	Ala 285	Tyr	Asn	His
20	Tyr	Ala 290	Gly	Pro	Lys	Glu	Ile 295	Arg	Ile	Tyr	Pro	Tyr 300	Asn	Asn	His	Glu
	Gly 305	Gly	Gly	Ser	Phe	Gln 310	Ala	Ile	Glu	Gln	Val 315	Lys	Phe	Leu	Lys	Arg 320
25	Leu	Phe	Glu	Glu	Gly 325											

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 298 amino acids
 - (B) TYPE: amino acid

30

35

40

45

50

55

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Met Arg Thr Leu Ser Phe Gly Pro Met Thr Thr Gly Gly Ser Ile His Met Ala Thr Met Asp Val Met Arg Gly Pro Gly Met Gln Arg Leu Ser 20 25 Gln Gly Ala Arg Glu Ala Ala Asn His Pro Trp Ala Lys Arg Leu Gly 35 40 Arg Met Gly Tyr Ala Ala Lys Gly Ala Val Tyr Ala Ile Ile Gly Val 50 55 60 Leu Ala Leu Lys Leu Ala Ala Gly Glu Gly Gly Arg Thr Thr Asp Ser 65 70 75 His Gly Ala Val Asn Thr Val Ala His Gly Pro Phe Gly Val Ala Leu 85 90 Leu Ala Val Leu Val Val Gly Leu Leu Gly Tyr Val Val Trp Arg Phe 105 Ala Gln Ala Phe Val Asp Thr Glu Asp Lys Gly Ser Asp Ala Lys Gly 115 125

	Ile	Ala 130	Thr	Arg	Ala	Met	Tyr 135	Phe	Leu	Ser	Gly	Cys 140	Ile	Tyr	Ala	Ser
5	Leu 145	Ala	Phe	Phe	Ala	Ala 150	Gln	Ser	Leu	Val	Gly 155	Ala	Ala	His	Gly	Arg 160
	Ser	Lys	Gly	Thr	Gln 165	Gly	Trp	Thr	Ala	Thr 170	Leu	Met	Glu	Gln	Pro 175	Phe
	Gly	Arg	Val	Leu 180	Val	Ala	Leu	Val	Gly 185	Leu	Gly	Ile	Val	Gly 190	Phe	Ala
10	Leu	Lys	Gln 195	Phe	His	Thr	Ala	Trp 200	Lys	Ala	Lys	Phe	Arg 205	Glu	Lys	Leu
	Thr	Leu 210	Thr	Gly	Leu	Ala	Ala 215	Arg	Lys	Gln	His	His 220	Ile	Glu	Arg	Met
	Cys 225	Gln	Phe	Gly	Ile	Ala 230	Ala	Arg	Gly	Val	Val 235	Phe	Ala	Val	Ile	Gly 240
	Gly	Phe	Leu	Val	Arg 245	Ser	Ala	Val	Asp	Ala 250	Asn	Pro	Gly	Glu	Ala 255	Lys
	Gly	Leu	Gly	Glu 260	Ala	Leu	Ala	Val	Val 265	Ala	Arg	Gln	Pro	Ser 270	Gly	Asp
20	Val	Leu	Leu 275	Gly	Val	Val	Ala	Ala 280	Gly	Leu	Val	Ala	Tyr 285	Ala	Ala	Tyr
	Leu	Phe 290	Leu	Gln	Ala	Arg	Tyr 295	Arg	Glu	Leu						

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 304 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Met Ser Lys Phe Ala Ile Leu Trp Ala Leu Ile Thr Ala Tyr Leu Pro Glu Pro Val Met Lys Leu Val Tyr Leu Gly Arg Arg Glu Thr Leu Gly Ala Arg Thr Leu Asp Val Lys Ala Gln Ala Val Gly Arg Leu Ala Asn Ala Thr Arg Pro Val Gly Val Ile Pro Thr Val Glu Glu Ser Arg Lys Met Thr Asp Lys Ala Val Ser Leu Phe Asp Gln Pro Ala Pro Glu Leu Phe Arg Lys Lys Asp Ile Gln Ile Asp Gly Ala Glu Gly Pro Ile Asp Ala Arg Ile Tyr Ser Gly Pro Ala Lys His Arg Xaa Xaa Pro Xaa Leu Val Tyr Phe His Gly Gly Gly Trp Val Gln Gly Asn Leu Asp Ser His

	Asp	Gly 130	Val	Cys	Gly	Lys	Leu 135	Ala	Lys	Trp	Ala	Asn 140	Cys	Ile	Val	Ile
5	Ser 145	Val	Asp	Tyr	Arg	Leu 150	Ala	Pro	Glu	His	Lys 155	Phe	Pro	Cys	Ala	Pro 160
	Leu	Asp	Ala	Ile	Ala 165	Ala	Tyr	Lys	Trp	Val 170	Arg	Ala	Asn	Ala	Thr 175	Asn
	Leu	Gly	Gly	Asp 180	Pro	Glu	Arg	Ile	Gly 185	Val	Gly	Gly	Asp	Ser 190	Ala	Gly
10	Gly	Asn	Leu 195	Ala	Ala	Val	Val	Cys 200	Gln	Gln	Thr	Ala	Met 205	Asn	Gly	Glu
	Arg	Thr 210	Pro	Asp	Leu	Gln	Val 215	Leu	Ile	Tyr	Pro	Ala 220	Leu	Asp	Ala	Arg
15	Met 225	Ile	Ser	Thr	Ser	Met 230	Glu	Glu	Leu	Arg	Asp 235	Ala	Tyr	Ile	Leu	Pro 240
	Lys	Ser	Arg	Met	Glu 245	Tyr	Phe	Leu	Gly	Leu 250	Tyr	Thr	Arg	Gly	Pro 255	Asp
	Asp	Ile	Glu	Asp 260	Leu	Arg	Met	Ser	Pro 265	Ile	Leu	Arg	Asp	Thr 270	Val	Ala
20	Asp	Gln	Pro 275	Gln	Ala	Cys	Ile	Val 280	Thr	Cys	Gly	Phe	Asp 285	Pro	Ala	Arg
· -	Arg	Arg 290	Glu	His	Leu	Arg	Arg 295	Thr	Leu	Asn	Cys	Arg 300	Gly	Asp	Arg	Arg

25

30

35

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 308 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

40

45

50

Va	l Ser 1	Ile	Arg	Leu 5	Arg	Leu	Leu	Asn	Trp 10	Phe	Leu	Asn	Thr	Phe 15	Glu
Lý	s Pro	Lys	Leu 20	Ala	Ala	Ala	Lys	Thr 25	Pro	Asp	Asp	Leu	Arg 30	Lys	Ser
Ph	e Glu	Leu 35	Lys	Ala	Arg	Phe	Leu 40	Phe	Pro	Ala	Pro	Arg 45	Lys	Thr	Arg
Ph	e Ser 50		Asp	Val	Leu	Gln 55	Ser	Gly	Ile	Gly	Ser 60	Val	Asn	Ala	Gln
Tr 6	p Ala 5	Lys	Ser	Lys	Ser 70	Ala	Ser	Asp	Asp	Arg 75	Val	Ile	Leu	Tyr	Phe 80
Hi	s Gly	Gly	Gly	Tyr 85	Val	Phe	Gly	Ser	Pro 90	Lys	Thr	His	Arg	Ala 95	Met
Le	u Ala	Arg	Leu 100	Ser	Ala	Met	Thr	Gly 105	Leu	Ser	Ala	Cys	Leu 110	Pro	Asp
Ту	r Arg	Leu 115	Ala	Pro	Glu	His	Pro 120	Phe	Pro	Ala	Ala	Ile 125	Glu	Asp	Ala
Va	1 Leu 130		Tyr	Lys	Cys	Leu 135	Leu	Glu	Arg	Ala	Ile 140	Glu	Pro	Gln	Asn

	Ile 145	Ile	Leu	Gly	Gly	Asp 150	Ser	Ala	Gly	Gly	Gly 155	Leu	Val	Leu	Ala	Leu 160
5	Leu	Ala	Glu	Ile	Lys 165	Ala	Gln	Ser	Leu	Pro 170	Lys	Pro	Ala	Gly	Val 175	Phe
	Ala	Leu	Ser	Pro 180	Leu	Val	Asp	Leu	Ser 185	Phe	Ser	Gly	Leu	Ser 190	Phe	Ser
	Lys	Asn	Ala 195	Gln	Thr	Asp	Val	Met 200	Leu	Pro	Ala	Ser	Arg 205	Ala	Ala	Asp
10	Met	Ala 210	Thr	Leu	Tyr	Leu	Asp 215	Gly	Ala	Asp	Ala	Asp 220	Asp	Pro	Arg	Ala
	Ser 225	Pro	Leu	Gln	Ala	Asp 230	Phe	Ser	Gly	Met	Pro 235	Pro	Val	Phe	Leu	Thr 240
15 ·	Ala	Ser	Asp	Ser	Glu 245	Ile	Leu	Leu	Asp	Asp 250	Cys	Leu	Arg	Met	Ala 255	Asp
	His	Leu	Arg	Ala 260	Gln	Gly	Val	Val	Val 265	Thr	Asp	Arg	Ile	Val 270	Glu	Asn
	His	Pro	His 275	Val	Trp	His	Ile	Phe 280	Gln	Arg.	Leu	Leu	Pro 285	Glu	Ala	Asp
20	Gln	Gly 290	Leu	Arg	Ala	Ile	Ala 295	Ala	Trp	Ile	Lys	Pro 300	Leu	Leu	Ser	Gly
	Ser 305	Asn	Glu	Ser												

(2) INFORMATION FOR SEQ ID NO:60:

25

30

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 237 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

40	Met 1	Leu	Thr	Phe	Asn 5	Val	Leu	Tyr	Gly	Met 10	Met	Lys	Gln	Lys	Leu 15	Ala	
	Ala	Ile	Leu	Met 20	Phe	Leu	Gly	Leu	Ser 25	Ala	Ala	Glu	Ala	Gln 30	Asp	Trp	
45	Pro	Asp	Leu 35	Gln	Lys	Tyr	Arg	Ser 40	Ala	Asn	Lys	Glu	Ala 45	Lys	Leu	Leu	
	Pro	Lys 50	Glu	Asn	Arg	Lys	Val 55	Val	Phe	Met	Gly	Asn 60	Ser	İle	Thr	Glu	
	Ala 65	Trp	Ile	Ser	Gln	Arg 70	Pro	Glu	Phe	Phe	Ser 75	Glu	Asn	Gly	Phe	Ile 80	
50	Gly	Arg	Gly	Ile	Ser 85	Gly	Gln	Thr	Thr	Pro 90	Gln	Met	Leu	Leu	Arg 95	Phe	
	Arg	Gln	Asp	Val- 100	Ile	Asp	Leu	Gln	Pro 105	Lys	Ala	Val	Val	Ile 110	Leu	Ala	
	Gly	Thr	Asn 115	Asp	Val	Ala	Gln	Asn 120	Thr	Gly	Pro	Met	Thr 125	Ile	Glu	Glu	
55	Ser	Leu 130	Ala	Asn	Ile	Lys	Ser 135	Met	Val	Glu	Leu	Ala	Gln	Ala	Asn	Gly	

Ile Thr Pro Val Leu Cys Thr Val Leu Pro Ala Asp Arg Phe Ser Trp 150 145 155 Arg Pro Glu Leu Thr Pro Ala Glu Thr Ile Ile Ala Leu Asn Gln Leu 165 170 Ile Lys Gln Tyr Ala Glu Ala Gln Gly Leu Ala Leu Val Asp Tyr His 180 185 190 Ala Ala Leu Thr Asn Lys Gly Gly Gly Leu Pro Val Lys Tyr Gly Glu 195 200 205 Asp Gly Val His Pro Asn Val Ala Gly Tyr Gln Val Met Glu Asn Ile 210 215 220 Val Leu Pro Val Ile Ser Ser Glu Leu Ala Lys Leu Lys 225 230

(2) INFORMATION FOR SEQ ID NO:61:

10

15

20

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 325 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

30	Met 1	Ala	Phe	Phe	Asp 5	Leu	Pro	Leu	Glu	Glu 10	Leu	Lys	Lys	Tyr	Arg 15	Pro
	Glu	Arg	Tyr	Glu 20	Glu	Lys	Asp	Phe	Asp 25	Glu	Phe	Trp	Glu	Glu 30	Thr	Leu
35	Ala	Glu	Ser 35	Glu	Lys	Phe	Pro	Leu 40	Asp	Pro	Val	Phe	Glu 45	Arg	Met	Glu
ω	Ser	His 50	Leu	Lys	Thr	Val	Glu 55	Ala	Tyr	Asp	Val	Thr 60	Phe	Ser	Gly	Tyr
	Arg 65	Gly	Gln	Arg	Ile	Lys 70	Gly	Trp	Leu	Leu	Val 75	Pro	Lys	Leu	Glu	Glu 80
40 ·	Glu	Lys	Leu	Pro	Cys 85	Val	Val	Gln	Tyr	Ile 90	Gly	Tyr	Asn	Gly	Gly 95	Arg
	Gly	Phe	Pro	His 100	Asp	Trp	Leu	Phe	Trp 105	Pro	Ser	Met	Gly	Tyr 110	Ile	Cys
		Val	115					120			_	-	125	_	_	-
45	Thr	Pro 130	Asp	Туr	Pro	Glu	Gly 135	Pro	Val	Asp	Pro	Gln 140	Tyr	Pro	Gly	Phe
	Met 145	Thr	Arg	Gly	Ile	Leu 150	Asp	Pro	Arg	Thr	Tyr 155	Tyr	Tyr	Arg	Arg	Val 160
		Thr	_		165					170					175	
50	Val	Asp	Gln	Glu 180	Arg	Ile	Val	Ile	Ala 185	Gly	Gly	Ser	Gln	Gly 190	Gly	Gly
	Ile	Ala	Leu 195	Ala	Val	Ser	Ala	Leu 200	Ser	Lys	Lys	Ala	Lys 205	Ala	Leu	Leu
55	Cys	Asp 210	Val	Pro	Phe	Leu	Cys 215	His	Phe	Arg	Arg	Ala 220	Val	Gln	Leu	Val
	Asp	Thr	His	Pro	Tyr	Ala	Glu	Ile	Thr	Asn	Phe	Leu	Lys	Thr	His	Arg

	225					230					235					240
•	Asp	Lys	Glu	Glu	Ile 245	Val	Phe	Arg	Thr	Leu 250	Ser	Tyr	Phe	Asp	Gly 255	·Val
	Asn	Phe	Ala	Ala 260	Arg	Ala	Lys	Ile	Pro 265	Ala	Leu	Phe.	Ser	Val 270	Gly	Leu
	Met	Asp	Asn 275	Ile	Cys	Pro	Pro	Ser 280	Thr	Val	Phe	Ala	Ala 285	Tyr	Asn	Tyr
	Tyr	Ala 290	Gly	Pro	Lys	Glu	Ile 295	Arg	Ile	Tyr	Pro	Tyr 300	Asn	Asn	His	Glu
	Gly 305	Gly	Gly	Ser	Phe	Gln 310	Ala	Val	Glu	Gln	Val 315	Lys	Phe	Leu	Lys	Lys 320
	Leu	Phe	Glu	Lys	Gly 325											•

(2) INFORMATION FOR SEQ ID NO:62:

10

15

25

30

35

45

50

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 210 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Leu 1	Lys	Tyr	Phe	Lys 5	Ala	Arg	Leu	Ala	Gly 10	Ile	Thr	Leu	Leu	Gly 15	Leu
Leu	Ala	Cys	Thr 20	Ser	Ala	Ser	Ala	Gln 25	Thr	Glu	Pro	Ile	Val 30	Phe	Val
His	Gly	Tyr 35	Ser	Gly	Ser	Ala	Ser 40	Asn	Trp	Asp	Thr	Met 45	Leu	Gly	Arg
Phe	Arg 50	Ser	Asn	Gly	Tyr	Ala 55	Ser	Gly	Ser	Leu	Tyr 60	Thr	Phe	Asn	Tyr
Asn 65	Ser	Leu	Val	Ser	Ser 70	Asn	Arg	Thr	Ser	Ala 75	Ser	Glu	Leu	Arg	Ser 80
Phe	Val	Asn	Thr	Val 85	Arg	Ser	Arg	His	Gly 90	Asn	Ala	Arg	Ile	Ala 95	Leu
Val	Ala	His	Ser 100	Asn	Gly	Gly	Leu	Val 105	Ser	Arg	Trp	Tyr	Arg 110	Ala	Glu
Leu	Gly	Gly 115	Glu	Thr	Ala	Thr	Arg 120	Arg	Phe	Val	Thr	Leu 125	Gly	Thr	Pro
His	Arg 130	Gly	Thr	Thr	Trp	Ala 135	Tyr	Ala	Cys	Tyr	Ser 140	Pro	Ala	Суѕ	Phe
Glu 145	Met	Arg	Pro	Gly	Ser 150	Ser	Leu	Leu	Thr	Thr 155	Leu	Gly	Ser	Arg	Ala 160
Cys	Asp	Arg	Ser	Leu 165	Trp	Ser	Asn	Thr	Asp 170	Gly	Ile	Ile	Leu	Pro 175	Ala
Ser	Ser	Ala	Gln 180	Cys	Gly	Val	Ser	Thr 185	Arg	Thr	Ala	Asp	Val 190	Ser	His
Leu	Asp	Leu 195	Leu	Thr	Asp	Ser	Arg 200	Val	Tyr	Thr	Gln	Leu 205	Arg	Thr	Gln
Leu	Gln 210											_			

Claims

5

10

15

20

30

35

40

45

50

55

- 1. A polynucleotide comprising a member selected from the group consisting of:
 - a) a polynucleotide comprising a nucleic acid sequence set forth in SEQ ID NO:23;
 - b) a polynucleotide which encodes an enzyme comprising an amino acid sequence set forth in SEQ ID NO:33;
 - c) a polynucleotide comprising at least 15 consecutive bases of a polynucleotide of (a) or (b);
 - d) a polynucleotide having at least 70%, 80%, 90%, 95% or 97% identity to the polynucleotide of (a) or (b); and encoding an enzyme having esterase activity; and
 - e) a polynucleotide which is complementary to a polynucleotide of any one of (a) to (d).
- 2. A polynucleotide hybridizing under stringent conditions to:
 - a) a polynucleotide comprising a nucleic acid sequence set forth in SEQ ID NO:23; or
 - b) a polynucleotide which is complementary to a polynucleotide of (a);

and encoding an enzyme having esterase activity.

- 3. The polynucleotide of claim 1 or 2, wherein the polynucleotide is DNA or RNA.
- 4. A vector comprising the polynucleotide of any one of claims 1 to 3.
- 5. A host cell comprising the vector of claim 4.
- 6. A process for producing a polypeptide comprising expressing from the host cell of claim 5 a polypeptide encoded by said polynucleotide.
 - A process for producing a cell comprising transforming or transfecting the cell with the vector of claim 4 such that the cell expresses the polypeptide encoded by the polynucleotide contained in the vector.
 - 8. An enzyme comprising a member selected from the group consisting of:
 - (a) an enzyme encoded by the polynucleotide of any one of claims 1(a), (b) or (d);
 - (b) an enzyme comprising an amino acid sequence which is at least 70%, 90% or 95% identical to an amino acid sequence set forth in SEQ ID NO:33;
 - (c) an enzyme which comprises at least 30 amino acids of the enzyme o (a) or (b); and
 - (d) an enzyme obtainable by the process of claim 6.
 - 9. An antibody specifically recognizing the enzyme of claim 8.
 - 10. A method for transferring an amino group from an amino acid to an α-keto acid comprising contacting an amino acid in the presence of an α-keto acid with the enzyme of claim 8.
 - 11. A method for the isolation of the enzyme of claim 8 comprising using the antibody of claim 9.
 - 12. Use of the enzyme of claim 8:
 - (a) as ripening starter for cheese in the dairy industry;
 - (b) for lignin removal in the pulp and paper industry;
 - (c) for the synthesis of carbohydrate derivatives; and/or
 - (d) for the conversion of lignocellulosic wastes to fermentable sugars.
 - 13. Use of the polynucleotide of claim 1 or 2 for the identification of similar sequences in other organisms.

Patentansprüche

1. Polynucleotid, umfassend ein Mitglied ausgewählt aus der Gruppe bestehend aus:

- (a) einem Polynucleotid, umfassend eine Nucleinsäuresequenz wie in SEQ ID NR:23 dargestellt;
- (b) einem Polynucleotid, das ein Enzym codiert, umfassend eine Aminosäuresequenz wie in SEQ ID NR:33 dargestellt;
- (c) einem Polynucleotid, umfassend mindestens 15 aufeinanderfolgende Basen eines Polynucleotids aus (a) oder (b);
- (d) einem Polynucleotid, das mindestens 70%, 80%, 90%, 95% oder 97% Identität mit dem Polynucleotid aus
- (a) oder (b) hat; und ein Enzym mit Esterase-Aktivität codiert; und
- (e) einem Polynucleotid, das zu einem Polynucleotid aus (a) bis (d) komplementär ist.
- 2. Polynucleotid, das unter stringenten Bedingungen an:

5

15

30

35

50

55

- (a) ein Polynucleotid, umfassend eine Nucleinsäuresequenz wie in SEQ ID NR:23 dargestellt; oder
- (b) ein Polynucleotid, das zu einem Polynucleotid aus (a) komplementär ist; hybridisiert und ein Enzym mit Esterase-Aktivität codiert.
- 3. Polynucleotid nach Anspruch 1 oder 2, wobei das Polynucleotid DNA oder RNA ist.
- 4. Vektor, umfassend das Polynucleotid nach einem der Ansprüche 1 bis 3.
- 20 5. Wirtszelle, umfassend den Vektor nach Anspruch 4.
 - Verfahren zur Herstellung eines Polypeptids, umfassend das Exprimieren eines Polypeptids, das von dem Polynucleotid codiert wird, von der Wirtszelle nach Anspruch 5.
- Verfahren zur Herstellung einer Zelle, umfassend das Transformieren oder Transfizieren der Zelle mit dem Vektor nach Anspruch 4, so dass die Zelle das Polypeptid exprimiert, das von dem im Vektor enthaltenen Polynucleotid codiert wird.
 - 8. Enzym, umfassend ein Mitglied ausgewählt aus der Gruppe bestehend aus:
 - (a) einem Enzym, das von dem Polynucleotid nach einem der Ansprüche 1 (a), (b) oder (d) codiert wird;
 - (b) einem Enzym, umfassend eine Aminosäuresequenz, die zu mindestens 70%, 90% oder 95% identisch ist zu einer Aminosäuresequenz wie in SEQ ID NR:33 dargestellt;
 - (c) einem Enzym, das mindestens 30 Aminosäuren des Enzyms aus (a) oder (b) umfasst; und
 - (d) einem Enzym, erhältlich durch das Verfahren nach Anspruch 6.
 - 9. Antikörper, der spezifisch das Enzym nach Anspruch 8 erkennt.
- Verfahren zum Transferieren einer Aminogruppe von einer Aminosäure zu einer α-Ketosäure, umfassend das
 Inkontaktbringen einer Aminosäure mit dem Enzym nach Anspruch 8 in Anwesenheit einer α-Ketosäure.
 - Verfahren zum Isolieren des Enzyms nach Anspruch 8, umfassend die Verwendung des Antik\u00f6rpers nach Anspruch 9.
- 45 12. Verwendung des Enzyms nach Anspruch 8:
 - (a) als Reifestarter für Käse in der Milchindustrie;
 - (b) zur Ligninentfernung in der Zellstoffindustrie;
 - (c) zur Synthese von Kohlenhydratderivaten; und/oder
 - (d) zur Umwandlung von Lignozelluloseabfällen zu fermentierbaren Zuckern.
 - Verwendung des Polynucleotids nach Anspruch 1 oder 2 zur Identifizierung von ähnlichen Sequenzen in anderen Organismen.

Revendications

1. Polynucléotide comprenant un membre choisi dans le groupe constitué de :

- a) un polynucléotide comprenant une séquence d'acide nucléique représentée dans SEQ ID NO : 23 :
- b) un polynucléotide qui code une enzyme comprenant une séquence d'acides aminés représentée dans SEQ ID NO : 33 ;
- c) un polynucléotide comprenant au moins 15 bases consécutives d'un polynucléotide de a) ou b);
- d) un polynucléotide ayant au moins 70%, 80%, 90%, 95% ou 97% d'identité avec le polynucléotide de a) ou
- b) ; et codant une enzyme ayant une activité estérase ; et
- e) un polynucleotide complémentaire de l'un quelconque des polynucléotides de a) à d).
- 2. Polynucléotide qui hybride dans des conditions stringentes à :
 - a) un polynucléotide comprenant une séquence d'acide nucléique représentée dans SEQ ID NO : 23 ; ou
 - b) un polynucléotide complémentaire d'un polynucléotide de a);

et codant une enzyme ayant une activité estérase.

15

10

- 3. Polynucléotide selon la revendication 1 ou 2, dans lequel le polynucléotide est de l' ADN ou de l'ARN.
- 4. Vecteur comprenant le polynucléotide selon l'une quelconque des revendications 1 à 3.
- 20 5. Cellule hôte comprenant le vecteur selon la revendication 4.
 - Procédé de production d'un polypeptide comprenant l'expression, à partir de la cellule hôte selon la revendication
 d'un polypeptide codé par ledit polynucléotide.
- 7. Procédé de production d'une cellule comprenant la transformation ou la transfection de la cellule avec le vecteur selon la revendication 4 de sorte que la cellule exprime le polypeptide codé par le polynucléotide contenu dans le vecteur.
 - 8. Enzyme comprenant un membre choisi dans le groupe constitué de :

30

35

40

- a) une enzyme codée par le polynucléotide selon l'une quelconque des revendications 1(a), (b) ou (d);
- b) une enzyme comprenant une séquence d'acides aminés identique à au moins 70%, 90% ou 95% à une séquence d'acides aminés représentée dans SEQ ID NO : 33 ;
- c) une enzyme comprenant au moins 30 acides aminés de l'enzyme de (a) ou (b) ; et
- d) une enzyme susceptible d'être obtenue par le procédé de la revendication 6.
- 9. Anticorps reconnaissant spécifiquement l'enzyme selon la revendication 8.
- 10. Méthode pour transférer un groupe amino d'un acide aminé à un α -céto acide, comprenant la mise en contact d'un acide aminé en présence d'un α -céto acide avec l'enzyme selon la revendication 8.
 - Méthode pour isoler l'enzyme selon la revendication 8, comprenant l'utilisation de l'anticorps selon la revendication
 9.
- 12. Utilisation de l'enzyme selon la revendication 8 :
 - a) comme inducteur de la maturation du fromage dans l'industrie laitière ;
 - b) pour éliminer la lignine dans l'industrie de la pâte et du papier ;
 - c) pour la synthèse de dérivés d'hydrates de carbone ; et/ou
 - d) pour la conversion de déchets ligno-cellulosiques en sucres fermentables.
 - 13. Utilisation du polynucléotide de la revendication 1 ou 2 pour l'identification de séquences similaires dans d'autres organismes.

55

FIGURE 1

Staphylothermus marinus - F1-1210

ATG	101	117	AAC	: AAG	CYC	101	TGG	ATG	GY1	ATG	ATA	ATA	TI	ATT	. CLC	AGC	. 111	, ICI	TTC	6
мес	ser	. rer	ABD	Lys	H19	ser	TEP	Met	Asp	Met	Ile	Ile	Phe	Ile	Leu	Ser	Phe	Ser	Phe	. 20
CCV	TTA	ACA	ATG	ATC	GCA	TTA	GCT	ATC	707	ATO	TCG	TCA	TGG	177	AAT	ATA	TGG	AAT	AAT	120
PIO	Leu	TAX	Met	ITE	Ala	Leu	Ala	Ile	Ser	Het	Ser	Ser	Trp	Phe	Asn	Ile	IIP	Авл	Asn	. 40
GCA	TTA	AGC	GAT	CTA	GGA	CAT	GCT	GII	AAA	AGC	AGT	GIT	GCT	, cox	ATA	TIC	AAT	CIA	GGT	180
Aļa	Leu	ser	Asp	Leu	Gly	H19	Ala	Val	Lys	Ser	Ser	Val	Ala	Pro	Ile	Phe	Aen	Leu	Gly	60
CII	GCA	ATT	COT	CCC	ATA	CTA	ATT	GIT	ATA	GTI	GGT	TTA	AGA	AAT	CIT	TAT	703	TGG	AGT	240
Leu	Ala	Ile	Gly	CIA	He	Leu	He	Val	Ile	Val	Gly	Leu	Arg	Asn	Leu	Tyr	Ser	IIP	Ser	80
AGA	GII	AAA	GGA	TCT	TTA	ATC	ATA	TCC	ATG	GGT	GTA	TIT	CII	AAC	TTA	ATA	GGG	GII	TTC	3,00
Arg	Val	Lya	gly	Ser	Leu	Ile	Ile	Ser	Met	Gly	Val	Phe	Leu	Asn	Leu	Ile	Gly	Val	Phe	100
GAC	GAA	GTA	TAT	GGT	TGG	ATA	CAT	TTC	CTA	GTC	TCA	GTA	TTG	TIT	TIC	TTA	TCA	ATA	ATA	360
Asp	Glu	Val	Tyr	Gly	dıı	Ile	His	Phe	Leu	Val	Ser	Val	Leu	Phe	Phe	Leu	Ser	Ile	lle	120
							ATA													420
IJΒ	Tyr	Phe	Ile	Ala	Ile	Ser	Ile	Leu	Asp	Lys	Sex	Trp	Ile	Ala	Val	Leu	Leu	Ile	Ile	140
							CTA													480
lly	His	He	YIS	Met	ттр	TYY	Leu	H18	Phe	Ala	ser	Glu	Ile	Pro	Arg	Gly	Ala	Ala	Ile	160
:00	G/G	TTA	TIA	aca	GTA	TTC	TCG	111	TTA	CCA	TTC	TAT	ATA	AGA	CAC	TAT	III	AAA	TCA	540
LO	GT.R.	Leu	rea	WTØ	val	rne	Ser	rne	ren	Pro	PRE	TYE	ITE	Arg	ASP	TYT	rne	Lye	ser	180
			CGA	TAG																555
yr	Thr	TAR	AFG																	185

FIGURE 2

Pyrodictium - TAG11-17LC

AT Me	g ar t Ly	a C	rc (eu j		GA:	Pro	C AC	A AA' F AB	T AC	C TO	C TA	C AC	T Le	G TI	EA C	re en	T T	TA eu	AT 9	Lei	G CA	T 60
II	r GC	A 17	rr 1	TAC	TGO	711	CIX	3 GC1	e GIY	TA	r ac	a TG	G 71	a CC	e a	T G7	e c	TA :	GTC	· ca	: cc	C 120
Pho	: AL	a Pt	ie 7	.YE	Irp	Phe	Let	ı Ali	ı Val	Ty	r Thi	r IX	p Le	u Pr	o G)	y Va	1 L	BU '	Val	AI	3 GJ	y .40
GT7 Val	AL	t GT a Va	1 A	ap ap	ACA Thr	GTA GGG	Val	GCT Ala	CGG Arg	GI(Va.)	Pro	01; co	y Le	n 81 c ee	y Ax	a yr g ca	g G	ly l	aag Lys	AGG	J Lei	1 60 1 60
CTC	CT	3 60	CG	CT	GTG	GCI	OTO	110	GCG	CTI	877	GT	TC	c gr	T ()1	TOI	ca	30 9	3CT	TAI	. CI	3 240
									Ala													
A)	Ty	r ag	T A	er Gr	Leu	His	bro CCG	GT n	AGC Sex	CAS	Arg	Pro	Va.	I AL	o cc	o gy g gw	3 (3C)	y I	eu.	ACC	Tyr	300 100
aaa Lys	GJ/	TT Ph	C A	GC (GTG Val	ACC Thx	GCG Ala	GAG Glu	GAT Asp	GCC	Leu	GTG Val	GT:	r yzi	3 GG 3 G1	C TG	3 G1 9 Va	3 6	TG æu	GGC Gly	Pro	360 120
GC	GCT	GGG	GG	CA	AC (:CG (GTG	TTC	GTT	TTG .	ATG	CAC	aca	TAT	ACT	CCC	TGC	~	20 1	NYI .	GCY:	120
Gly	Ale	9 G1	y G	ly i	Asn	Pro	Val	Phe	Val	Leu	Met	His	Gly	יעד י	Th	r Gl	Cy	e A	rg	Ser	ALA	140
ccc	TAC	: AT	G	CT (GTG	CJG	GCC	CGG	GAG	CIC	GIG	COAG	7 G0	GGG	TA	2 cc	GI	G G	TT	GTG	TTC	480
									Glu				-	-	_							
GAC Asp	Phe	CG	g Gi	iy i	CAC Hib	gj Å geg	GJ n GVO	AGC Ser	CJA CCG	GGC	TCG Ser	ACG	ACC	Ile	GL	Pro	CG Ar	g G	AG lu	GTG Val	CTG Leu	540 180
GAT	GCC	: CG(3 G(T (376	GTG	GGC	TAT	GTC	TCO	GAO	CGG	770	: ccc	: GG(: 000	: CG	3 A	TA	ATA	TIG	600
									Val													200
a	GGG	II	K	T !	ATG	GGC	GGC	CCI	GTA	ecc	ATC	GIG	GAG	GGT	GC	GGG	GA	c ō	CG	CGG	GTC	660
									Val							_						220
TAC Tyr	GCG Ala	Val	GC	IT (SCT Ala	GAT Asp	AGC Ser	Pro	TAC Tyr	TAI	AGG Arg	Leu	CGG	GAC	Val	: ATA	Pro	C CA	GG Tg	TGG TTP	ren CIG	720 240
C N CT	Th/	330		· ·	~~		~~	~~~	TGG	~~	com	~~		~~~		-	-	. ~	~	200	~	780
Glu	Tyr	Lys	Th	r i	20	Leu	Pro	Gly	Trp	Va1	Gly	Val	Leu	Ala	Gly	Phe	Ty	G	ly i	Arg	Leu	260
									GGC													840
Het	Ala	Gly	· Va	l A	sp 1	leu	Gly	Phe	Gly	Pro	Ala	Gly	Val	Glu	Arg	Val	Asį	L	/16 J	Pro	Leu	280
									Pro													900 ·
rec	CGT	AGC	co	G T	GT (GC ·	CGT	CTC	GTC	GAG	OTT	CCI	GGG	GC7	GGC	CAC	GTG	GEA	0	3CC	GTG	960
er	Arg	Ser	Pr	0 C	ув (aly i	Arg	Leu	Val	Glu	Val	Pro	GJA	Ala	Gly	His	Val	Q1	.u /	lla '	Val	320
									GCA (1020
-						_	_	.yr	Ala	rap	me C	weu	TIG	OTG	nea	wrg	478	G1	.	, LU	Cys	340
		GJ À GGG					TGA															1041 306

PIGURE 3

Archaeoglobus Venificus SN P6-24LC

ATX Hot	Pro	TA:	r GT	r ago	AA! ABI	GG1	GTA GCI	GTA Val	AA1	ATC	TA:	TA	CA)	Leu Leu	G1G Val	GAT Asp	. G(1)	CC!	GAG Glu	. 6 2
Pro	CCF Pro	AT.	GT(Phe	GIT Val	CAC His	Gly	TGG	ACA Thr	GC: Ala	AA:	T ATC	AA1	TII Phe	11b	Lys	GAC Glu	G)1	AGA Arg	12 4
																			Ser Ser	18
GAT Asp	Lys	Pro	CTI Leu	GLY	Tyr	GAT Asp	TTC Phe	TAC Tyr	AGA Arg	771 Phe	GAC Glu	AAC Asn	Phe	ATT Ile	TCA Ser	CAT Asp	TTA Leu	GAT Asp	GCG	246
GTT Val	GII Val	AGG	GJu GAG	ACT	GLY	GTG Val	GJ n GYG	aaa Lyb	TTT Phe	GII Val	CTC	GTC Val	GCA	CAT His	TCA Ser	TTC Phe	GGA Gly	ACA Thr	ATG	300 100
														GCT Ala						360 120
G17 GG1	GCG Gly	AGC Ser	AGA Arg	ATA Ile	aag Lyb	CTT	CTA Leu	CAC His	AGA Arg	ATT Ile	GIY	TAT Tyr	CCI Pro	TTA Leu	GCA Ala	aag Lyb	ATT Ile	CTT Leu	GCA Ala	420 140
														TCC Ser.						480 160
														ACA Thr						540 180
let	Tyr	Thr	Tyr	Arg	Thr	Leu	Thr	Lys	Val	Asn	Leu	Glu	Asn	ATC Ile	Leu	Glu	Lya	Ile	Asp	600 200
:ya	Pro	Thr	Leu	Ile	Ile	Val	Gly	Glu	Glu	Asp	ALA	Leu	Leu	Pro	Val:	Ser :	Lys	9er	Val	660 220
llu	Leu	Ser	Arg	Arg	Ile	Glu	Asn .	Ser :	Lys	Leu	Val	Ile	Ile	PTO	Asn :	Ser	giy	His	CAB	720 240
/al	Met	Leu												GJu i						780 260
in.	TTC Phe	IAA																		789 263

PIGURE 4

Aquifex pyrophilus - 28LC

TI	S AGU	TI	AGC	Z AAZ	TI	GN	GAC	ATA	AAC	CTC	G	CI	TCC	GGA	GGA	GC1	CC	AAG	GGC	6
LG	ı arı	, rec	. Arg	, mye	PRE	: GIU	ı Giu	r ire	ABI	, ren	. VA.	. Leu	5er	GTA	GTA	ALL	Ale	Lys	Gly	. 20
ATZ	GCC	: CAC	ATA:	GGI	GIT	110	AAA :	GCT	ATA	AAC	CAG	CTC	GGT	ATA	AGG	GIG	AGG	GCT	TTA	120
116	: Ala	H76	11e	GIY	. APT	. Leu	Lys	Ala	Ile	: Asn	Glu	Leu	Gly	Ile	Arg	Val	Arg	Ala	Leu	. '40
AGC	GGG	GIG	AGC	GCC	GGG	CCA	ATC	GTT	TCG	GTC	777	TAT	GCC	TCA	GGC	TAC	TCC	CCT	GAA	180
Ser	GIA	Val	Ser	Ala	Gly	Ala	Ile	Val	Ser	Val	Phe	Tyr	Ala	Ser	Gly	LÀI	Ser	Pro	Glu	60
CCC	ATG	TTC	AGC	CII	CIG	AAG	AGG	GTA	AAC	TGG	CTG	AAG	CIG	TTT	AAG	TTC	AAG	CCA	CCT	240
Gly	Met	Phe	Ser	Leu	Leu	Lys	Arg	Val	Asn	TIP	Leu	Lys	Leu	Phe	Lys	Phe	Lys	Pro	Pro	80
CTG	AAG	GGA	TTG	ATA	GGG	TGG	GAG	AAG	GCT	ATA	AGA	TTC	CTT	GAG	GAA	GIT	CTC	CCT	TAC	300
Leu	Lys	Gly	Leu	Ile	GJA	Trp	Glu	Lys	Ala	lle	Arg	Phe	Leu	Glu	Glu	Val	Leu	Pro	Tyr	100
agc	AGA	ATA	COAA	AAA	CII	GAG	ATA	CCG	ACG	TAT	λΤλ	TGC	GCG	ACO	CAT	TTA	TAC	TCG	GGA	360
Arg	Arg	lle	Glu	Lys	Leu	Glu	Ile	Pro	Thr	Tyr	11e	Сув	Ala	Thr	Asp	Leu	Tyr	Ser	Gly	120
AGG	CCT	CTA	TAC	CIC	TCG	GAA	GGG	AGT	TIA	ATC	ccc	GCA	CTT	CIC	ccc	PUC	Tit	GCJV	ATT	420
Arg	Ala	Leu	TYE	Leu	Ser	Glu	Gly	Ser	Leu	lle	Pro	Ala	Leu	Leu	Gly	Ser	Сув	Ala	Ile	140
ccc	GGC	ATA	TIT	GAA	ecc	GTT	GAG	TAT	AAG	AAT	TAC	TIG	CTC	GIT	GAC"	GGA	CCT	ATA	CTT	480
bro	gly	Ile	Phe	Glu	Pro	Val	Glu	Tyr	Lys	Asn	Tyx	Leu	Leu	Val	Asp	Gly	Gly	Ile	Val	160
AAC	AAC	CII	ccc	GIT	GAG	222	TIT	CAG	CAA	AGC	GGT	ATT	ccc	ACC -	GTT	TOP	GTT	CAT	ctrc	540
Asn	Asn	Leu	Pro	Val	Glu	Pro	Phe	Gln	Glu	Ser	gly	Ile	Pro	Thr	Val	Сув	Val	Авр	Val	180
CTT	ccc	ATA	GAG	cca	GAA	AAG	GAT	ATA	DAA	AAC -	ATT		CAC	ATC:	TT	THY:	a/C/C	AGC	-	600
Leu	Pro	Ile	Glu	Pro	Glu	Lys	Asp	lle	Lys	ABD	Ile	Leu	Hib	Ile	Leu	Leu	Arg	Ser	Phe	200
777	CIT	GCG	310	ccc	TCA	AAC	TCC	GAA	ANG	AGA	ANG	- 	TIT	TGT (GAC		GTT	ATA	CTT.	660
Phe	Leu	Ala	Val	Arg	ser	Asn	Ser	GJu	Lys	Arg	Lys	Glu .	Phe	CAR	Asp :	Leu	Val	Ile	Val	220
133	GAG	CII	GAG	CAG	TTC	ACA	ccc	CTT (CAT	GTT .	AGA	AAA	GCG (GAC (CAA :	ATA :	DTG	CAC:	NC3	72ò
PTO	Glu	Leu	Glu	Glu	Phe	Thr	Pro	Leu	Asp	Val	Arg	Lys	Ala :	Aap (מונ	Lle	Met	Glu i	Arg	240
EQA.	TAC	ATA	AAG	CCC	TTA	GAG (GTA (CII :	ici (GAA '	TAG									756
	Tyr																			252

FIGURE 5

MIITL-29L.

AT Me	G II	T AA e As	T A	TC ; le ;	TAA RBA	GT(Phe	Val	TAA 1	r ATA	A TC	TIG	CIC	TA:	r TT:	r TC	A GG	G AI	A GI e Va	T ATG 1 Met	60 20
																				T GTA	_
· Ly	B Th	r Va	1 0	lu (31u	Tyr	Al	Lei	Lev	Gli	Th	Gly	Va)	Arg	Val	l Phe	Ty:	r CG	g Cy	r GTA s Val	120
																			_	C AGT	
îi	e Pr	0 01	u L	א פץ	ũa	Phe	A9E	Thx	Leu	116	ile	Gly	Ser	. His	Gly	Let	(01)	/ Ala	a Hi	C AGT B Ser	180 60
GG	A AT	C TA	C A	T A	GT	GII	GCI	GAA	GAA	TIT	GCT	AGG	CAC	GGA	. 777	GGZ	170	TG	. ATY	CAC	240
G1	y 11	e Ty	r II	le S	er	Val	Ala	GJ n	Glu	Phe	Ala	Arg	His	Gly	Phe	Gly	Phe	CY	Met	His	BO
GA	T CA	A AG	G G	A C	AT	GGG	AGA	ACG	GCA	AGC	CAT	ACA	GAA	AGA	GGG	TAT	GTG	CAC	GGG	: 111	300
As	p GL	תגו מ	g G1	ун	18	GIA	Arg	Thr	Ala	Sex	Asp	Arg	Glu	Arg	Gly	Tyx	Val	. Glu	(31)	Phe	100
CA	C AA	: 11	CA	D A	AG	GAT	ATG	AAG	GCC	TTC	TCC	GAT	TAT	GCC	DAA	TOG	CGC	GTG	GGJ	GOT	360
Hi	s Ası	2 Ph	e II	e G	lu	ABP	Met	Lys	Ala	Phe	Ser	Asp	TYI	Ala	Lys	<u>1</u> cb	Arg	Va]	Gly	Gly	120
GA	c cas	AT	IA &	A T	76	CTA	GGA	CAC	AGT	ATG	GGC	GGG	CIG	ATA	GCG	CTC	TTA	ACA	GII	CCA	420
Ası	Glu	ille	Il	e L	eu	Leu	Gly	His	Ser	Het	Gly	gjy	Leu	Ile	Ala	Leu	Leu	Thr	Val	Ala	140
AC1	TAT	AAJ	CA	A A	TC	GCC	AAG	GGA	GIT	ATC	GCG	CTA	GCC	CCG	GCC	ctc	CAA	ATC	ccc	TTA	480
Thi	Tyr	Lys	a1	u I.	le	Ala	Lys	Gly	Val	Ile	Ala	Leu	Ala	Pro	Ala	Leu	Gln	Ile	Pro	Leu	160
ACC	cea	GCT	AG	A AC	3A	CTT	GIT	CTA	AGC	CTC	GCG	TCA	agg	CIT	GCC	cca	CAT	TCT	AAG	ATC	540
Thr	Pro	Ala	Ar	g Aı	rg .	Leu	Val	Leu	Ser	Leu	Ala	Ser	Arg	Leu	Ala	Pro	His	Ser	Lys	Ile	180
ACC	TTA	CAR	AG	3 AG	2A :	TTG	cce	CAG	AAA	CCA	GAG	GGT	TIT	CAA	AGA	GCA	AAA	GAT	ATA	GAA	600
Thr	Leu	gin	Ar	y Ar	9	ren	PYO	GIR	Lys	Pro	Glu	gry	Phe	Gln	Arg	Ala	Lye	Asp	He	Glu	200
	AGT																				660
TYE	Ser	Leu	361	. 41	u .	116	ser	ANT	гув	Leu	Val	АВР	01 <i>a</i>	met	IIG	rys	A18	ser	Ser	MOE	220
	TOO																				720 240
	_					_											-		-	-	240
	GTC Val																				780 260
												-								Ξ,	
	TTG																				840 280
-		-		_			-		•								•			-,-	:
	GTC Val																	TAR			894 298
								•		•											

PIGURS 6

Thermococcus CL-2-30LC

																			CAT BLR	· 60
																			GCC	120 40
	GGA Gly																		CAC	180
ACG	AGC	GTC	DAD	GAC	GCG	ATG	AAD	ATC	ATC	GAC	700	ATA	ATC	GAG	GAG	ATC	AGG	GAG	AAG	240
	Sex									•							_		•	.80
	TTC Phe																			300 100
	GAT ABP																			360 120
	GGC																			420
	AAC							•			. •	•				•				140
	Asn																			480 160
	yab Gyc																			540 180
	GAG Glu																			600 200
	GGC Gly																			660 220
	AAC Asn																			720 240
	GCC Ala																			780: 260;
	CAA										-									789 263

FIGURE 7

Aquifex VP5-34LC

																			A GGG	
					_		•	_										-	A AAG	_
Ala	Ala	Ly	GL	, Ile	Ala	His	Ile	Gly	Val	Leu	Lys	Ala	Les	Glu	GJU	Lev	Gly	, Ile	Lys	120
GTA	AAG	AGG	CTC	AGC	GGG	GE	AGT	GCT	GGA	GCI	ATC	GI	100	GTC	777	TAC	: cc	TCC	GCC	190
									_							_			01y	
Tyt	Thx	Pro	Asp Asp	GJu	Met	Leu	Lys	ren	Leu	Lys	GAG Glu	GTA Val	AAC -Ast	Trp	Leu	Lys	Leu	Phe	Lys	240 80
																			AAA	300
Phe	Lys	Thi	Pro	Lys	Met	Gly	Leu	Met	GIY	Lib	G)11	Lys	Ala	Ala	Glu	Phe	Leu	Glu	Lys	, 100
							GAA Glu													360 120
							TAC													420
	- • -	-	•	•			Tyr		•	•	•	•					_		-•	240
							TIT Phe													480 160
							CCC Pro													540 180
							ACT													600
Gly	Val	Asp	Val	Leu	Pro	Ile	Thr	Gln	Glu	Arg	Lys	Ile	Lys	Asn	Ile	Leu	HīB	Ile	Leu	200
							Val													660 220
							CAA .													720
							GJu			SEL	PTO	reu	мар	ABT	ASR	₽γγ	MIB	vab	ata	240
							GCA Ala		IAA											750 250

PTOTER

Teredinibacter - 42L

at Me	g cc t Pr	A GC	AA T Bab	T GA	C TC	A CC	C ACI	ATC	GAC Ast	TT	T AA' a Abi	T CC	T CG	g G1	C AT	T CT	r cc	C AA g As	C GC	F 60
CA:	C GC	A CAG	GT Va	r at	r TT	A GC	G ACT	TCC	GCC	Tro	CG(C AA	A GO	7 TT 2 Pb	e Le	n Tài	A CG	C AC	G CAC F His	2 120 4 40
AA(Lys	AG(TAC	CTC	AG(ACT	OO 1	CAA	Trp	CIG	G) (CTC	C GA:	C GCC	GG G1;	C AA	c GGI	GT Va	Th:	C TTC	180
GC(Ala	GC	G) U	Let	AA(ABI	The	GCC	CCT Pro	GCA Ala	ACT Thr	GCA Ala	TCC Ser	TCC Ser	TCC Sex	CAC Hid	CCC	GCC Ala	CA Hi:	Lyi	S AAC S ASI	240 80
ACT Thr	Leu	GTT Val	AI7	Val	Leu	CAC His	GGC	TGG	GAA Glu	G1A GGC	TCC Ser	AGC Sei	Gl:	TCC Sea	GCC Ale	TAT	GCC	ACC	TCC Sex	300 100
A) a	GC	AGC Ser	ACC	Leu	TTC	GAC Asp	AAT Aan	GGG Gly	TTC Phe	GAC	ACT	Phe	CGC	Lev	CAA ?	Phe	Arg CG(GA7	CAC His	360 120
GGC	GAC Asp	Thr	TAC	His	Leu	AAC	CGC Arg	GCC	ATA Ile	TII Phe	AAC Asn	TCA Ser	TCG Ser	Leu	ATI	CAC GAC	Glu	Val	GTO Val	420 140
Gly	Ala	Val	Lya	Ala	Ile	Gl n	G]n	Gln	Thr	Ysb	Tyr	Asp	Lys	Tyr	Сув	Leu	Het	Gly		480 160
Ser	Leu	Gly	Gly	Asn	Phe	Ala	Leu	Arg	Val	Ala	Val	Arg	Glu	GJu	His	Leu	Ala	Lys		540 180
ren	Ala	Gly	Val	Leu	Ala	Val	Cys	Pro	Val	Leu	увр	Pxo	Ala	His	Thr	Met	Met	Ala	Leu	200 200
Asn	Arg	Gly	Ala	Phe	Phe	Tyr	GCC	Arg	TYT	Phe	Ala	His	Lys	Trp	Lys	Arg	Ser	Leu	Thr	660 220
Ala	Lye	Leu	Ala	Ala	Phe	Pro	GAC Asp	Tyr	Lys	Tyr	<u> </u>	Lys	Asp	Leu	Lys	Ser	Ile	His	Thr	720 240
Leu	Asp	Glu	Leu	Aon	Авп	TYX	TTC Phe	Ile	Pro	Arg	Tyr	Thr	Gly	Phe	Asn	Ser	Val	Ser	Glu	780 260
Tyr	Phe	Lys	Sex	Tyr	Thr	Leu	ACC	Gly	Gln	Lys	Leu	Ala	Phe	Leu	Asn	Сув	Pro	Ser	Tyr	940 280
lle	Leu	Ala	Ala	Gly	Asp	уер	Pro	Ile	Ile	Pro	Ala	Ser	Авр	Phe	Gln	Lys	Ile	Ala	Lys	900 300:
Pro	Ala	Agn	Leu	His	Ile	Thr	GTA Val	Thr (Gln (Gln	GJA	Ser	Hie	CAR	Ala	Tyr	Leu	Сјп		960 320
							gac Asp													339

FIGURE 9

Archeoglobus fulgidas VC16 - 16MC1

ATG CIT GAT ATG CCA ATC GAC CCT GTT TAC TAC CAG CIT GCT GAG TAT Met Leu Asp Met Pro Ile Asp Pro Val Tyr Tyr Gln Leu Ala Glu Tyr TTC GAC AGT CTG CCG AAG TTC GAC CAG TTT TCC TCG GCC AGA GAG TAC Phe Asp Ser Leu Pro Lys Phe Asp Gln Phe Ser Ser Ala Arg Glu Tyr agg gag gcg ata aat cga ata tac gag gag aga aac cgg cag ctg agc Arg Glu Ala Ile Asn Arg Ile Tyr Glu Glu Arg Asn Arg Gln Leu Ser CAG CAT GAG AGG GTT GAA AGA GTT GAG GAC AGG ACG ATT AAG GGG AGG Gln His Glu Arg Val Glu Arg Val Glu Asp Arg Thr Ile Lys Gly Arg 55 AAC GGA GAC ATC AGA GTC AGA GTT TAC CAG CAG AAG CCC GAT TCC CCG Asn Gly Asp Ile Arg Val Arg Val Tyr Gln Gln Lys Pro Asp Ser Pro GGT CTG GTT TAC TAT CAC GGT GGT GGA TTT GTG ATT TGC AGC ATC GAG Val Leu Val Tyr Tyr His Gly Gly Gly Phe Val Ile Cys Ser Ile Glu tcg cac gac gcc tta tgc agg aga ayy gcg aga ctt tca aac tct acc Ser His Asp Ala Leu Cys Arg Arg Ile Ala Arg Leu Ser Asn Ser Thr GTA GTC TCC GTG GAT TAC AGG CTC GCT CCT GAG CAC AAG TTT CCC CCC Val Val Ser Val Asp Tyr Arg Leu Ala Pro Glu His Lys Phe Pro Ala cea git tat cat tgc tac gat gcg acc aag tgg git gct gag aac cgg Ala Val Tyr Asp Cys Tyr Asp Ala Thr Lys Trp Val Ala Glu Asn Ala GAG GAG CTG AGG ATT GAC CCG TCA AAA ATC TTC GTT GGG GGG GAC AGT Glu Glu Leu Arg Ile Asp Pro Ser Lys Ile Phe Val Gly Gly Asp Ser GCG GGA CGG AAT CTT GCC CCG GCG CTT TCA ATA ATG GCG AGA GAC AGC Ala Gly Gly Asn Leu Ala Ala Ala Val Ser Ile Met Ala Arg Asp Ser gga gaa gat tic ata aag cat caa att cta act tac ccc git gig aac Gly Glu Asp Phe Ile Lys His Gln Ile Leu Ile Tyr Pro Val Val Asn 185 TIT GIA GCC CCC ACA CCA TCG CIT CTG GAG TIT GGA GAG GGG CTG TGG Phe Val Ala Pro Thr Pro Ser Leu Leu Glu Phe Gly Glu Gly Leu Trp att ctc gac cag aag ata atg agt tgg ttc tcg gag cag tac ttc tcc Ile Leu Asp Gln Lys Ile Met Ser Trp Phe Ser Glu Gln Tyr Phe Ser

AGA GAG GAA GAT AAG TTC AAG CCC CTC GCC TCC GTA ATC TTT GCG GAC Arg Glu Glu Aso Lys Phe Asn Pro Leu Ala Ser Val Ile Phe Ala Asp

245

CTT GAG AAC CTA CCT CCT GCG CTG ATC ATA ACC GCC GAA TAC GAC CCG Leu Glu Asn Leu Pro Pro Ala Leu Ile Ile Thr Ala Glu Tyr Asp Pro 255 260 265

CTG AGA GAT GAA GGA GAA GTT TTC GGG CAG ATG CTG AGA AGA GCC GGT Leu Arg Asp Glu Gly Glu Val Phe Gly Gln Met Leu Arg Arg Ala Gly 270 280

GTT GAG GCG AGC ATC GTC AGA TAC AGA GGC GTG CTT CAC GGA TTC ATC Val Glu Ala Ser Ile Val Arg Tyr Arg Gly Val Leu His Gly Phe Ile 285

AAT TAC TAT CCC GTG CTG AAG GCT GCG AGG GAT GCG ATA AAC CAG ATT ASN Tyr Tyr Pro Val Leu Lys Ala Ala Arg Asp Ala Ile Asn Gln Ile 300 305 310

GCC GCT CTT CTT GTG TTC GAC TAG Ala Ala Leu Leu Val Phe Asp 315 320

FIGURE 10

Sulfolobus Solfatarious P1 - 8LC1

ATG CCC CTA GAT CCT AGA ATT AAA AAG TTA CTA GAA TCA GCT CTT ACT Met Pro Leu Asp Pro Arg Ile Lys Lys Leu Leu Glu Ser Ala Leu Thr ATA CCA ATT GGT AAA GCC CCA GTA GAA GAG GTA AGA AAG ATA TTT AGG Ile Pro Ile Gly Lys Ala Pro Val Glu Glu Val Arg Lys Ile Phe Arg CAA TTA GCG TCG GCA GCT CCC AAA GTC GAA GTT GGA AAA GTA GAA GAT Gin Leu Ala Ser Ala Ala Pro Lys Val Glu Val Gly Lys Val Glu Asp ata aaa ata cca ggc agt gaa acc git ata aac gct aga gig tat tit Ile Lys Ile Pro Gly Ser Glu Thr Val Ile Asm Ala Arg Val Tyr Phe CCG AAG AGT AGC GGT CCT TAT GGT GTT CTA GTG TAT CTT CAT GGA GGC Pro Lys Ser Ser Gly Pro Tyr Gly Val Leu Val Tyr Leu His Gly Gly GGT TIT GTA ATA GGC GAT GTG GAA TCT TAT GAC CCA TTA TGT AGA GCA Gly Phe Val Ile Gly Asp Val Glu Ser Tyr Asp Pro Leu Cys Arg Ala ATT ACA AAT GCG TGC AAT TGC GTT GTA GTA TCA GTG GAC TAT AGG TTA lle Thr Asn Ala Cys Asn Cys Val Val Val Ser Val Asp Tyr Arg Leu GCT CCA GAA TAC AAG TIT CCT TCT GCA GTT ATC GAT TCA TIT GAC GCT Ala Pro Glu Tyr Lys Phe Pro Ser Ala Val Ile Asp Ser Phe Asp Ala 120 act aat tgg gtt tat aac aat tta gat aaa ttt gat gga aag atg gga Thr Asn Trp Val Tyr Asn Asn Leu Asp Lys Phe Asp Gly Lys Met Gly GTT GCG ATT GCG GGA GAT AGT GCT GGA GGA AAT TTG GCA GCG GTT GTA Val Ala Ile Ala Gly Asp Ser Ale Gly Gly Asn Leu Ala Ala Val Val 155 GCT CTT CTT TCA AAG GGT AAA ATT AAT TIG AAG TAT CAA ATA CTG GTT Ala Leu Leu Ser Lys Gly Lys Ile Asn Leu Lys Tyr Gln Ile Leu Val TAC CCA GCG GTA AGT TTA GAT AAC GTT TCA AGA TCC ATG ATA GAG TAC Tyr Pro Ala Val Ser Leu Asp Asn Val Ser Arg Ser Met Ile Glu Tyr TCT GAT GGG TTC TTC CTT ACC AGA GAG CAT ATA GAG TGG TTC GGT TCT Ser Asp Gly Phe Phe Leu Thr Arg Glu His Ile Glu Trp Phe Gly Ser

200

CAA TAC TTA CGA AGC CCT GCA GAT TTG CTA GAC TTT AGG TTC TCT CCA Gln Tyr Leu Arg Ser Pro Ala Asp Leu Leu Asp Phe Arg Phe Ser Pro

220

195

ATT CTG GCG CAA GAT TTC AAC GGA TTA CCT CCA GCC TTG ATA ATA ACA Ile Leu Ala Gln Asp Phe Asn Gly Leu Pro Pro Ala Leu Ile Ile Thr 225 230 235 240

GCA GAA TAC GAT CCA CTA AGG GAT CAA GGA GAA GCG TAT GCA AAT AAA Ala Glu Tyr Asp Pro Leu Arg Asp Gln Gly Glu Ala Tyr Ala Asn Lys 245 250 255

CTA CTA CAA GCT GGA GTC TCA GTT ACT AGT GTG AGA TTT AAC AAC GTT Leu Leu Gln Ala Gly Val Ser Val Thr Ser Val Arg Phe Asn Asn Val 260 265 270

ATA CAC GGA TTC CTC TCA TTC TTT CCG TTG ATG GAG CAA GGA AGA GAT Ile His Gly Phe Leu Ser Phe Phe Pro Leu Met Glu Gln Gly Arg Asp 275

GCT ATA GGT CTG ATA GGG TCT GTG TTA AGA CGA GTA TTT TAT GAT AAA Ala Ile Gly Leu Ile Gly Ser Val Leu Arg Arg Val Phe Tyr Asp Lys 290 295 300

ATT TAA

Ile

305

Pigore 11 LA11.1 Esterase es23

ATG	AAG	077	. AAA	CAC	GII	ATT	GII	TTA	CAT	GGC	TTA	TAT	DTA	TCI	GGC	TTG	GTC	ATG	CCC	. 60
Mat	Lye	Val	Lys	Hig	Val	Ile	Val	Leu	His	Gly	Leu	Tyr	Met	Ser	gly	Leu	Val	Met	Arg	20
~~				~~	~~	~~~	C10	~~~	-	~~						. ~	-			
																			Thr	: 120
PIU	, DEG	Сув	361	My	<i>D</i> eu	GTU	GIU	Ser	GIY	AgT	ьув	ANT	Leu	ABII	Lett	int	ıyı	ABI	THE	40
COA	GAC	CCT	TAA	CGA	GAT	GCT	ATT	TII	ACG	CAA	ATA	GAT	GAG	III	ATT	AGC	AAT	GAG	CCT	180
																			Pro	60
~	-	-	~~~	-	~~	~~	1.00	~~~	^~	-	~	~~	~~~	~~~				~~*	***	240
																			AAC ABD	80
361	WIG	100	Val	Cyb	nab	Set	HEL	GIY	GIY	Den	VAL	Wid	My	MI	TAL	Ten	GIB	MIG	MBG	, 80
TCA	GCG	CCA	AGT	CAT	CAT	CII	GAA	AAG	GTA	ATC	ACC	TTA	GGA	ACG	CCA	CAT	ACT	GGC	AGC	300
Ser	Ala	Pro	Ser	His	His	Va1	Glu	Lys	Val	Ile	Thr	Leu	QJA	The	SLO	- His	Thr	Gly	Ser	100
						~~	~		~~~		~~	~							010	250
																		GII Val		360 120
HIS	110	YIA	GLU	гув	Het	GID	OIN	nys	GIA	PER	Gru	ren	Den	Trên.	гув	ABII	set	val	GIU	120
TIT	TTA	CTC	TCT	AAG	AAT	GGT	GAT	TGG	CCT	TIT	AAA	GCC	AAG	CTA	TAT	AGC	ATT	GCC	GCC	420
Phe	Leu	Leu	Ser	Lys	Asn	Oly	qeA	Trp	Pro	Phe	Lys	Ala	Lys	Leu	Tyr	Ser	Ile	Ala	GJA	140
0 h C	777	~~	877	ccc	TTA	2.772	CC		እ ፕተ	GT3	222	aac	אמר	mar.	70-	(D) T	aac	ACT	GTA	480
																		Thr		160
-OP	~.			 ,								,		9			,			7.0
																		AGC		540
Leu	Leu	Asp	G1 tr	Thr	Lys	Leu	Lys	Gly	Met	YJ9	Olu	His	Lys	Val	Phe	alH	Leu	Ser	His	180
a ("A	AGT	ATG	ATT	TAC	767	ccc	CAA	atc	att	AAT	TAT	ATT	CTT	ന്മദ	CGC	TTG	AAC.	GAG	GAC	600
																		Glu		200
				- • -			-							-	•				•	
ATT	TA																			605
(le																				201

Pigure 12

Whale Mat Sample 11.801 Beterase es?

ATC	ATA	AAA	AAC	TIC	GAC	AG/	GA	AAT	101	AGC	177	GEA	CTC	TCC	GGT	GG	GG	r GC	י כזק	6
mec	116	rys	ABII	PRE	АВр	Arg	Gin	АБЛ	ser	ser	Leu	Val	Leu	ser	GIY	, GI	, G1 ³	Ala	Leu	3
GGT	ATT	CCI	CYC	TTO	GGT	GIA	C10	CAT	GAC	CIT	GY1	AAA	CAA	AAT	ATT	GIA	CO	AAT	GAA.	12
gtà	116	Ala	HIS	Leu	GTA	, vai	Leu	HIB	Asp	Leu	Glu	Lys	Gln	Aen	Ile	. Val	Pro	A91	Glu	. 4
ATT	GIT	GGT	ACA	AGT	ATG	001	CCI	ATC	ATT	GGT	GCA	TCI	ATG	GCT	ATC	GGG	ATG	. AAA	CAG	18(
ITe	val	gry	IDI	Ser	Met	gry	GIA	Ile	Ile	gīÀ	Ala	Ser	Met	λla	He	Gly	Met	Lys	Glu	61
AAA	GAA	ATA	CIC	GAA	GAA	ATC	AAA	AAC	TIT	TCC	AAT	GTC	TTC	AAC	TGG	ATA	AAA	110	TCT	240
Lys	Glu	Ile	Leu	Glu	Glu	Ile	Lya	aeA	Phe	Ser	Asn	Val	Phe	Asn	Trp	Ile	Lys	Pbe	Ser	80
III	TCC	GGT	AAT	TCT	GTT	GTC	GAT	AAC	GAG	DAA	ATC	GCT	AAG	ATA	TTT	GAT	ACT	CIT	TIT	300
Phe	Ser	Gly	Asn	Ser	Val	Val	yab	Asn	Glu	Lys	Ile	Ala	Lys	Ile	Phe	Asp	Thr	Leu	Phe	100
AAA	GAC	AGA	AAG	ATG	ACA	GAT	ACG	GTG	ATC	CCT	CII	AAA	CTC	ATC	GCT	ACA	AAC	TTA	CAT	360
Lya	qaA	Arg	Lys	Met	Thr	qaA	Thr	Val	Ile	Pro	Leu	Lys	Leu	Ile	Ala	Thr	Aen	Leu	His	120
AAT	GGA	CAT	AAA	AAA	GTA	TIT	ACT	GCT	TCG	CAT	GAT	GTA	CIG	ATC	AAA	CAT	GCA	ATA	CTC	420
as a	Gly	His	Lys	Lys	Val	Phe	Thr	Ala	Ser	Asp	Asp	Val	Leu	Ile	Lys	λsp	Ala	Ile	Leu	140
TCA	ACA	ATG	GCA	ATA	ccc	GGT	GTA	TTT	GAA	GAA	CAT	ATT	ATT	GAT	GGT	GAA	ACC	TAT	GGC	480
Ser	Thr	Met	Ala	Ile	PTO	GJA	Val	Phe	Q1u	Glu	His	lle	Ile	Asp	Gly	Glu	Thr	Tyr	Gly	160
CAC	GGT	TIT	CIT	TGT	GAA	AAC	CIT	GGT	GTG	AAT	GAG	GCA	ACA	TTC	TAK	GAT	GTT	TTA	GCT	540
rab	Gly	Phe	Leu	Cys	GJn	neA	Leu	Gly	Val	ABŊ	Glu	YJa	Thr	Phe	Asn	Asp	Val	Leu	Ala	180
JTA	gat	GTC	DTA	GGT	GAG	AAC	TCT	TIT	GAA	AAA	GCA	ATG	cca	GAC	AAC	TTC	TIT	AAA	ACA	600
Val	qaA	Val	Met	Gly	Glu	asa	Ser	Phe	Glu	Lys	6LA	Met	Pro	Asp	Asn	Phe	Phe	Lys	Thr	200
rca	AAT	GIT	TTA	GAA	ATG	TTT	GAA	AAA	TCA	DTA	CGA.	CIT	TIT	ATT	TAC	AAC	CAG	ACA	CAG	660
ser	Asn	Val	Leu	Glu	Met	Phe	Glu	Lys	Ser	Met .	Arg	Leu	Phe	Ile	Tyr	asa	Gln	Thr	Gl'n	220
								AAT .												720
Chr	His	Ile	Lys .	Asn .	Ala	Asn	Lys	Asn	Ile	IJI.	Leu	Ile	Glu	Pro	Val	Thr	Lys	Glu	Tyr	240
LAA	ACA	TIT	CAA	TTT .	CAT	AAA	CAT	AAA (GAG .	ATA	CGT	GCT '	TIA :	GGC	DII	GGT	TTA	CTG	TG	779
ys	Thr	Phe	Gln .	Pbe .	aiH	Lys	His	Lys (Glu	Ile /	Arg .	Ala :	Leu	Gly :	Leu ·	GJA	Leu	Leu		25,9

Pigure 13

Metallosphaera Prunae Ron 12/2 Esterase 23mcl

	ATG	CC	כ כד	A CA	$x \propto$	AA A	C CT	A AM	3 AA	A TT	A CT	T TC	C CY	G CT	A CCT	CCC	CA	o con	C TT	c rec	6
	Met	Pr	o Le	n HI	B Pr	о Гу	e va	l Ly	3 PA	B Lei	i Lei	r Se	r Gli	n Lei	1 SEC) bxc) GL	n Asi	P Pb	e Ser	2
	7(2)	AA	· GT	ദാമ	. GZ	· —	a Ma	2 22/	2 00	r TC	2 (23)		. ~	-	• •	oc.	3.72		2 624	g ACC	
	Arq	Ası	ı Va	1 G1	n As	o Le	u Arv	Lv	AL	a Tri	ο Ann	Le	n CC	o Phi	Ser	alv	Arc	i Are	Gl	u The	130
	CTG	AA	3 AG	g GT	T GA	G GA	c cm	CAL	AT:	A CCC	: AC	r AG	3 GA	: GCZ	CGA	ATC	AGG	GCC	: AG	gre	186
	Leu	Ly	. Ar	g Va	1 Gl	u AB)	p Lev	ı Glı	ı Ile	PEC	מעב כ	. Ar	g Asy	o Ala	. Arg	Ile	Arc	, Ali	Ar	y Val	60
	TAC	אמי	٠ ~		א או	P 331	2 (23.2	* **	• •	~~		. ~~		-	-	~~				TTC	
	īvr	Th	Pr	o Se	r Sei	LV	Gly	AAT		Dry	. Val	l La	· Val	1700	. 1/41	His	Giv	. 61.	(3)	, Phe	240 80
	-,-					,.				• •••	, ,		. 74.2	,.	,.	****	· (L.)	, G1,	QI.	7 4110	. 01
	GIG	TIC	: GG	r ag	C GT	CA(C AGG	TAC	: GA(: G00	: CIC	: GC2	1700	: CII	ATT	GCC	AAG	CAN	TCT	CCC	300
	Val	Phe	Gl	/ Se	r Val	L ASI) Ser	Tyx	Ası	gly	Lev	Ale	Sex	Lev	Ile	Ala	Lys	Gly	Ser	. GJA	100
		~~	. ~~		^ -~	-~															
																				Val	360 120
								-,-	3	, 200	-				2,2				744		
																				GAC	420
	λen	qaA	Sex	. Lel	Ast	Ala	Leu	Leu	Trp	Ile	Ala	Glu	מפג	GIA	gly	Lys	Leu	Gly	Leu	Asp	140
	•	-~	201			· CTC		~~~	03.4		~	~~	~~~		~			~	~~~	TCC.	480
					Ala																160
			_					_	_			•	-								200
																				GCA	540
	Leu	Leu	Asp	Arg	Asp	GID	GTA	Lys	gly	Leu	Val	Ser	IAL	GID	AgT	Leu	He	Tyr	Pro	Ala	180
	ата	226	ATY	GTC	GAI	AAC	TCC	~	TCC	atc	D/2/I	G) G	TAC	aac	CAG	ADD	TAC	TTC	CTC	204	600
					Aap																300
											-		•	•							'
					AAC																660
	Arg	Ser	wer	HOC	Aan	ith	Mie	gry	THE	Ret	TYT	Pne	ser	ser	шy	wd	GIU	WIG	AUT	Ser	220
	ccc	TAC	GCC	TCT	CCA	GCC	TTG	GCT	CAC	CTA	CAT	AAC	CTC	CCA	CCC	TCA	CIG	GTG	ATC	ACT	720
	Pro	TYT	Ala	Ser	PTO	Ala	Leu	Ala	Авр	Leu	Rio	Asn	Leu	Pro	bro	Ser	Leu	Val	Ile	Thr	240
	~~	C N C	~>~	ChT	ccc	CT.	NOC:	ON TO	ma	cen	620	۰		~~	mc.	**	~~	220	ana	-	780
					Pro																260
•			-,-							,			-/-	~							
					YCC																840
•	ala a	ASD:	val	ser	Thr	ren	ATT	ATG	TYT	GID	gly	Ket	116	H18	GIA	rne .	ren	ser	rds	TYT	280
	ana '	raa	ATA	ACT	GCC	GGT	AAA	CTA	GCC	ATT	CAC	CAC	ATT	CT	GGG (GIT .	CTG	AGA	TCT	GTC	300
					Ala																300
						-	-														
	CTT '	TA																			905 301
4	Ŀeυ																				304

Piqure 14

Thermotoga neepolitana 5068 Esterase 56mc4

																			C GAG	
Va.	. Ala	Pho	e Ph	AB	Met	Pro	Leu	Glu	Glu	Leu	Lye	Lys	נעד	. Arg	Pro	Glu	Arg	TY:	r Glu	20
																			CTG	
Glu	Lys	ABJ	Phe	Asy	Ġlı	Phe	, Jxb	Arg	Glu	The	Len	Lye	Glv	ser	. GJA	(G)	Phe	Pro	Leu	1. 40
																			ACT	180
Asp	Pro	Val	Phe	Glu	Lys	Val	Asp	Pho	His	Leu	Lys	Thr	Val	Glu	Thr	Tyr	Yet	Val	Thr	. 60
TTC	TCT	GGA	TAC	AGG	GGG	CAG	AGA	ATA	AAG	GGC	TGG	CIT	CTI	GIT	cca	AAG	TTG	GCG	CAA	. 240
							Arg													80
aan	מע	CTX	CCA	TGC	GTC	GTG	CAG	TAC	ATA	COT	TAC	AAT	GOT	(27)	700	GGT	717	CCA	CAC	300
							Gln													100
CNC	7000	CTO	***	T12/2	CCG	TCD	ATG	ocrr	TAC	22	***	77	GTC	בתים	CAC	200	A CO CO	aca	CRC	360
							Met													120
-		~~~	***	2007	220	~~~	GAC		~~	~ -			~~~	~~	~~	~~~	~~=	~~	m.o	420
							Asp													140
							-						-					~~~	~~~	
							GGC													480 160
- 4		•				•	•			•		•		•	•	-	•	•		•
							GTG													540 180
					_										_		-		- •	
							AGT Ser													600 200
•						-			-	-	-									
							CTC													660 220
		_		-				-	•					-			_	•		
							Pro													720 240
val	OTII	Leu	VAL	wab	1114	NIO	710	Lyr	vaı	G.U	116	THE	MBIA	FIRE	LEU.	Dyo	****	MIS	My	
							AGA Arg													780 260
кар	LYB	ota	GIG	110	Val	Mile	AL 9	1111	Dea	ser	IYE	*138	KUP	GIY	AUT	MOII	VIDE:	ALL	VIG	200
							TTT													840 280
arg	ALB	гAв	AST	MO	WIG	ren	Fire	ser	VAI	GIA	Leu	MUC	wab	IME	116	Cys	PLO	710	SET	400
							CAC													900
THE	Val	PRE	ATB	YIS	ıyı	ABD	Hie	Tyr .	ALE	GIA.	PTO	ну в	GTD	11e	Arg	116	lyt	PEO	1 y E	300
							TCT													960
ren	ASn	H18	GID	GIA	OIA	GIA	Ser	LD6 (GIR	WID .	TTG (OT#	GID	ABT	-ys	rne .	œu	υyσ	wa	320
			GAA		TAG															978 326
æu	rne	عالم	Glu	OIA																340

Pigure 15

Malittangium lichenicola Esterase 77mcl

ATG CGC ACC CTC TCC TTC GGT CGG ATG ACC ACA GGG GGA AGC ATT CAC ATG GGG ACC ATG Mat Arg Thr Leu Ser Phs Gly Pro Met Thr Thr Gly Gly Ser Ile Mis Met Ala Thr Me GAC GTO ATG CGC GGG CCG GGG ATG CAG CGG CTG TCA CAG GGC GCC AGG GAG GCC GGG AA AAP VAL Met Arg Gly Pro Gly Met Gln Arg Leu Ser Gln Gly Ala Arg Glu Ala Ala Ala App Val Met Arg Gly Pro Gly Met Gln Arg Leu Ser Gln Gly Ala Arg Glu Ala Ala Ala CAC CCC TGG GCG AAG CGA CTG GGC CGC ATG GGC TAC GCG GCC AAG GGC GCC GTG TAC GC His Pro Trp Ala Lys Arg Leu Gly Arg Met Gly Tyr Ala Ala Lys Gly Ala Val Tyr Ala ATC ATC GGC GTG CTC GCG CTG AAG CTC GCG GCG GGC GAG GGC GGG ACC ACG GAC Ile Ile Gly Val Leu Ala Leu Lys Leu Ala Ala Gly Glu Gly Gly Arg Thr Thr Anp Sei CAC GGC GCG GTG AAC ACC GTG GGG CAC CAG GCC CTTC GGC GTG CGC CTG GCG GTG GTG AAC ACC GGC GTG GAA ACC ACC GTG GCG CAC GGG CCC TTC GGC GTG CGC GTG GCG GTG GGC GTG GAAC ACC GTG GGC CAC GGG CCC TTC GGC GTG CGC CTG His Gly Ala Val Asm Thr Val Ala His Gly Pro Phe Gly Val Ala Leu Leu Ala Val Leu CTG GTG GGC CTG CTG GGC TAC GTG GTC TGG AGG TTC GCC CAG GCC TTC GTG GAC ACG Val Val Gly Leu Leu Gly Tyr Val Val Trp Arg Phe Ala Gln Ala Phe Val App Thr Glu GAC AAG GGC TCC GAC GGC AAG GGA ATC GCC ACG CGC GCC CAG GCC TTC GTG GAC ACG ABP Lys Gly Ser Asp Ala Lys Gly Ile Ala Thr Arg Ala Met Tyr Phe Leu Ser Gly Cys ACC AAG GGC TCC GAC GGC TTC GCC GCC CAG CCC TTC GTG GGC GCC GCC Ile Tyr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ris Gly Arg ACC AAG GGG TCG GAC GCC TTC GCC GCC CAG CCC GTG CTG GCC CTG GTG GGC TTG GCC TTC GTG GGC GCC GC	
ARE GITG ATG CGC GGG CCG GGG ATG CAG CGG CTG TTA CAG GGC GCC AGG GGC GGG AAAS VAL NAT ATG GLY PTO GLY MAT GIR ATG LEU SAT GLA GLY ALA ATG GLY ALA ALA AAS VAL NAT GLY PTO GLY MAT GIR ATG LEU SAT GLO GCC AAG GGC GCC GTG TAC GGC HIS PTO TTP ALA LYS ATG LEU GLY ATG MAT GGC GGC GAG GGC GAG GGC GGC GTG TAC GGC HIS PTO TTP ALA LYS ATG LEU GLY ATG MAT GGC GGC GAG GGC GAG GGC GGC GAG GGC GGC	3 60
ASP Val Not Arg Gly Pro Gly Not Gln Arg Leu Ser Gln Gly Ala Arg Glu Ala Ala Ala As CAC CCC TAG GCG AAG CGA CTG GGC CGC ATG GGC TAG GCC GCG GTG TAC GCC CCC TAG GCG GCG AAG GCC GCG GTG TAC GCC His Pro Trp Ala Lya Arg Leu Gly Arg Not Gly Tyr Ala Ala Lya Gly Ala Val Tyr Ala ATC ATC GGC GTG CTC GCG CTG AAG CTC GCG GCG GAG GGC GAG GGC GCC GGA ACC ACG GAC ACG Ile Ile Gly Val Leu Ala Leu Lya Leu Ala Ala Gly Glu Gly Gly Arg Thr Thr Ang Sei CAC GGC GCG GTG TAC ACC GGC GCG GTG AAC ACC GTG GCG CCC TTC GGC GTC GCG CTG CTG GCG GTG GT	5 20
CAC CCC TGG GCG AAG CGA CTG GGC CGC ATG GGC TAC GCG GCC AAG GGC GTG TAC GCG His Pro Trp Ala Lys Arg Leu Cly Arg Not Gly Tyr Ala Ala Lys Gly Ala Val Tyr Ala ATC ATC GGC GTG CTC GCG CTG AAG CTC GCG GCG GGC GAG GGC GGC CGG ACC ACG Ile Ile Gly Val Leu Ala Leu Lys Leu Ala Ala Gly Glu Gly Gly Arg Thr Thr Asp Sei CAC GGC GCG GTG AAC ACC GTG GCG CAC GGG CCC TTC GGC GTC GCG CTG CTG GCG GTG GT	120
His Pro Trp Ala Lys Arg Leu Gly Arg Net Gly Tyr Ala Ala Lys Gly Ala Val Tyr Ala NTC ATC GGC GTG CTC GCG CTG AAG CTC GCG GCG GGC GAG GGC GGC CGG ACC ACG GAC ACG Ile Ile Gly Val Leu Ala Leu Lys Leu Ala Ala Gly Glu Gly Gly Arg Thr Thr Asp Sei CAC GGC GCG GTG AAC ACC GTG GCG CAC GGG CCC TTC GGC GCG CCG CTG CTG GCG GTG CTG His Gly Ala Val Asm Thr Val Ala His Gly Pro Phe Gly Val Ala Leu Leu Ala Val Leu OTG GTG GGC CTG CTG GGC TAC GTG GTC TGG AGG TTC GCC CAG GCC TTC GTG GAC ACG GAG Val Val Gly Leu Leu Gly Tyr Val Val Trp Arg Phe Ala Gln Ala Phe Val Asp Thr Glu GAC AAG GGC TCC GAC GCG AAG GGA ATC GCC AAG CCC GCC ATG TAC TTC CTC AGC GGC TGC Asp Lys Gly Ser Asp Ala Lys Gly Ile Ala Thr Arg Ala Met Tyr Phe Leu Ser Gly Cys ATC TAC GCG TCG CTG GCC TTC TTC GCC GCG CAG TCC CTG GTG GGC GCG GCG CAC GCC CTG Lie Tyr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ris Gly Arg AGC AAG GGG ACG CAG GCC TGG ACG GCC ACG CTG ATG GAC CCC TTT GGC GGC GCG Ser Lys Gly Thr Gln Gly Trp Thr Ala Thr Leu Het Glu Gln Pro Phe Gly Arg Val Leu GTG GGG CTG GTG GGG CTG GGC ATC GTG GGC TTC GGG CTG AAG CAG TTC CAC ACC GGC Val Ala Leu Val Gly Leu Gly Ile Val Gly Phe Ala Leu Lys Gln Phe His Thr Ala Trp AAG GGG AAG TTC CGG GAG AAG CTC ACC CTC ACC GCA CTG GCT GCC CGG AAG CAC CAC Lys Ala Lys Phe Arg Glu Lys Leu Thr Leu Thr Gly Leu Ala Ala Arg Lys Gln His His ATC GAG CTC ATC TCC CCC GTG GAC ATC GTC GCC GGC GTG GTG TTC GCC CTG GAG CAC Lys Ala Lys Phe Arg Glu Lys Leu Thr Leu Thr Gly Leu Ala Ala Arg Lys Gln His His ATC GAG CTC GTC GTC CGC TTC GGC ATC GCC GCG GGC GTG GTG TTC GCC GTC ACG GCC TTC CTC GTC CGC TCC GCC GTG GAC GCG GCG GTG GTG TTC GCC GTG AAG GCC TTC CTC GTC CGC TCC GCC GTG GAC GCG GCG GTG GTG TTC GCC GTG GAG GCC TTC CTC GTC CGC TCC GCC GTG GAC GCG GCG GCG GTG GTG TTC GCC GCG Ala Leu Ala Val Val Ala Arg Gln Pro Ser Gly Asp Val Leu Leu Gly Val Val Val Ala Ala CGC CTG GTG GCC TAC GCC GCC GCC CTC GCC GCC GCC GCG GCG Ala Leu Ala Val Val Ala Arg Gln Pro Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GCC CTG GTG GCC TAC GCC	1 40
His Pro Trp Ala Lys Arg Leu Gly Arg Net Gly Tyr Ala Ala Lys Gly Ala Val Tyr Ala NTC ATC GGC GTG CTC GCG CTG AAG CTC GCG GCG GGC GAG GGC GGC CGG ACC ACG GAC ACG Ile Ile Gly Val Leu Ala Leu Lys Leu Ala Ala Gly Glu Gly Gly Arg Thr Thr Asp Sei CAC GGC GCG GTG AAC ACC GTG GCG CAC GGG CCC TTC GGC GCG CCG CTG CTG GCG GTG CTG His Gly Ala Val Asm Thr Val Ala His Gly Pro Phe Gly Val Ala Leu Leu Ala Val Leu OTG GTG GGC CTG CTG GGC TAC GTG GTC TGG AGG TTC GCC CAG GCC TTC GTG GAC ACG GAG Val Val Gly Leu Leu Gly Tyr Val Val Trp Arg Phe Ala Gln Ala Phe Val Asp Thr Glu GAC AAG GGC TCC GAC GCG AAG GGA ATC GCC AAG CCC GCC ATG TAC TTC CTC AGC GGC TGC Asp Lys Gly Ser Asp Ala Lys Gly Ile Ala Thr Arg Ala Met Tyr Phe Leu Ser Gly Cys ATC TAC GCG TCG CTG GCC TTC TTC GCC GCG CAG TCC CTG GTG GGC GCG GCG CAC GCC CTG Lie Tyr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ris Gly Arg AGC AAG GGG ACG CAG GCC TGG ACG GCC ACG CTG ATG GAC CCC TTT GGC GGC GCG Ser Lys Gly Thr Gln Gly Trp Thr Ala Thr Leu Het Glu Gln Pro Phe Gly Arg Val Leu GTG GGG CTG GTG GGG CTG GGC ATC GTG GGC TTC GGG CTG AAG CAG TTC CAC ACC GGC Val Ala Leu Val Gly Leu Gly Ile Val Gly Phe Ala Leu Lys Gln Phe His Thr Ala Trp AAG GGG AAG TTC CGG GAG AAG CTC ACC CTC ACC GCA CTG GCT GCC CGG AAG CAC CAC Lys Ala Lys Phe Arg Glu Lys Leu Thr Leu Thr Gly Leu Ala Ala Arg Lys Gln His His ATC GAG CTC ATC TCC CCC GTG GAC ATC GTC GCC GGC GTG GTG TTC GCC CTG GAG CAC Lys Ala Lys Phe Arg Glu Lys Leu Thr Leu Thr Gly Leu Ala Ala Arg Lys Gln His His ATC GAG CTC GTC GTC CGC TTC GGC ATC GCC GCG GGC GTG GTG TTC GCC GTC ACG GCC TTC CTC GTC CGC TCC GCC GTG GAC GCG GCG GTG GTG TTC GCC GTG AAG GCC TTC CTC GTC CGC TCC GCC GTG GAC GCG GCG GTG GTG TTC GCC GTG GAG GCC TTC CTC GTC CGC TCC GCC GTG GAC GCG GCG GCG GTG GTG TTC GCC GCG Ala Leu Ala Val Val Ala Arg Gln Pro Ser Gly Asp Val Leu Leu Gly Val Val Val Ala Ala CGC CTG GTG GCC TAC GCC GCC GCC CTC GCC GCC GCC GCG GCG Ala Leu Ala Val Val Ala Arg Gln Pro Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GCC CTG GTG GCC TAC GCC	2 180
CAC GGC GCG GTG AAC ACC GTG GCG CAC GGG CCC TTC GGC GTG CTG CTG GCG GTG CTG His Gly Als Val Asn Thr Val Als His Gly Pro Phe Gly Val Als Leu Leu Als Val Leu OTG GTG GGC CTG CTG GGC TAC GTG GTC TGG AGG TTC GCC CAG GCC TTC GTG CAC ACG GAG Val Val 3ly Leu Leu Gly Tyr Val Val Trp Arg Phe Als Gln Als Phe Val Asp Thr Glu GAC AAG GGC TCC GAC GCG AAG GGA ATC GCC ACG GCC ATG TAC TTC CTC AGC GGC TGC Asp Lys Gly Ser Asp Als Lys Gly Ile Als Thr Arg Als Net Tyr Phe Leu Ser Gly Cys ATC TAC GCG TCG CTG GCC TTC TTC GCC GCG CAG TCC CTG GTG GGC GCC GCG CAG Ile Tyr Als Ser Leu Als Phe Phe Als Als Gln Ser Leu Val Gly Als Als Ris Gly Arg ACC AAG GGG ACG CAG GGC ACG GCC ACG CTG ACG CAG CAG CAG CCC GTT TG GCC CGC Ile Tyr Als Ser Leu Als Phe Phe Als Als Gln Ser Leu Val Gly Als Als Ris Gly Arg ACC AAG GGG ACG CAG GGC TGG ACG GCC ACG CTG ATG GAG CAG CCC TTT GCC CGC GTG CTG Ser Lys Gly Thr Gln Gly Trp Thr Als Thr Leu Net Glu Gln Pro Phe Gly Arg Val Leu GTG GCG CTG GTG GGG CTG GGC ATC GTG GGC TTC CCG GTG AAG CAG TTC CAC ACC GGS TGG Val Als Leu Val Gly Leu Gly Ile Val Gly Phe Als Leu Lys Gln Phe His Thr Als Trp AAG GCG AAG TTC CGG GAG AAG CTC CTC ACC CGA CTG GCC GTG GTG CCC CGC GTG TGG ACC ATG TCC CAG TTC GGC ATC GCC GCC GGC GTG GTG TTC GCC ACC CAC CAC CAC GAG CCC ATG TGC CAG TTC GGC ATC GCC GGC GTG GTG TTC GCC GTC ATC GGC CAC CAC GAG CCC ATG TGC CAG TTC GGC ATC GCC GGC GTG GTG TTC GCC GTC ATC GGC CAC CAC GAG CCC ATG TGC CAG TTC GGC ATC GCC GGC GTG GTG TTC GCC GTC ATC GGC CAC GAG CCC ATG TGC CAG TTC GGC ATC GCC GGC GTG GTG GTG GCC GTG GTG CAC CTC GTC GCC CTC GCC GTG GAG CAG CCG GGC GTG GTG GTC GCC GGC Als Leu Als Val Als Arg Gln Phe Gly Ile Als Als Arg Gly Val Val Phe Als Val Ile Gly GGC TTC CTC GTC GCC GCC GCG GTG GTG GTG	60
CAC GGC GCG GTG AAC ACC GTG GCG CAC GGG CCC TTC GGC GTG CTG CTG GCG GTG CTG His Gly Als Val Asn Thr Val Als His Gly Pro Phe Gly Val Als Leu Leu Als Val Leu OTG GTG GGC CTG CTG GGC TAC GTG GTC TGG AGG TTC GCC CAG GCC TTC GTG CAC ACG GAG Val Val 3ly Leu Leu Gly Tyr Val Val Trp Arg Phe Als Gln Als Phe Val Asp Thr Glu GAC AAG GGC TCC GAC GCG AAG GGA ATC GCC ACG GCC ATG TAC TTC CTC AGC GGC TGC Asp Lys Gly Ser Asp Als Lys Gly Ile Als Thr Arg Als Net Tyr Phe Leu Ser Gly Cys ATC TAC GCG TCG CTG GCC TTC TTC GCC GCG CAG TCC CTG GTG GGC GCC GCG CAG Ile Tyr Als Ser Leu Als Phe Phe Als Als Gln Ser Leu Val Gly Als Als Ris Gly Arg ACC AAG GGG ACG CAG GGC ACG GCC ACG CTG ACG CAG CAG CAG CCC GTT TG GCC CGC Ile Tyr Als Ser Leu Als Phe Phe Als Als Gln Ser Leu Val Gly Als Als Ris Gly Arg ACC AAG GGG ACG CAG GGC TGG ACG GCC ACG CTG ATG GAG CAG CCC TTT GCC CGC GTG CTG Ser Lys Gly Thr Gln Gly Trp Thr Als Thr Leu Net Glu Gln Pro Phe Gly Arg Val Leu GTG GCG CTG GTG GGG CTG GGC ATC GTG GGC TTC CCG GTG AAG CAG TTC CAC ACC GGS TGG Val Als Leu Val Gly Leu Gly Ile Val Gly Phe Als Leu Lys Gln Phe His Thr Als Trp AAG GCG AAG TTC CGG GAG AAG CTC CTC ACC CGA CTG GCC GTG GTG CCC CGC GTG TGG ACC ATG TCC CAG TTC GGC ATC GCC GCC GGC GTG GTG TTC GCC ACC CAC CAC CAC GAG CCC ATG TGC CAG TTC GGC ATC GCC GGC GTG GTG TTC GCC GTC ATC GGC CAC CAC GAG CCC ATG TGC CAG TTC GGC ATC GCC GGC GTG GTG TTC GCC GTC ATC GGC CAC CAC GAG CCC ATG TGC CAG TTC GGC ATC GCC GGC GTG GTG TTC GCC GTC ATC GGC CAC GAG CCC ATG TGC CAG TTC GGC ATC GCC GGC GTG GTG GTG GCC GTG GTG CAC CTC GTC GCC CTC GCC GTG GAG CAG CCG GGC GTG GTG GTC GCC GGC Als Leu Als Val Als Arg Gln Phe Gly Ile Als Als Arg Gly Val Val Phe Als Val Ile Gly GGC TTC CTC GTC GCC GCC GCG GTG GTG GTG	240
His Gly Als Val Asn Thr Val Ala His Gly Pro Phe Gly Val Ala Leu Leu Ala Val Leu Otto GTO GGC CTG GGC TAC GTG GTC TGG AGG TTC GGC CAG GCC TTC GTG CAC ACG GAG Val Val Gly Leu Leu Gly Tyr Val Val Trp Arg Phe Ala Gln Ala Phe Val Asp Thr Glu CAC AAG GGC TCC GAC GCC AAG GGC TCC CAG GCC TCC CAG GGC CAG TCC CTG GTG GGC GGC CAG CAC GGC CAG ILe Tyr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ris Gly Arg AGC AAG GGG ACG GGC ACG GCC ACG CTG CTG GGC GGC CAG CAC CAC	80
His Gly Als Val Asn Thr Val Ala His Gly Pro Phe Gly Val Ala Leu Leu Ala Val Leu Otto GTO GGC CTG GGC TAC GTG GTC TGG AGG TTC GGC CAG GCC TTC GTG CAC ACG GAG Val Val Gly Leu Leu Gly Tyr Val Val Trp Arg Phe Ala Gln Ala Phe Val Asp Thr Glu CAC AAG GGC TCC GAC GCC AAG GGC TCC CAG GCC TCC CAG GGC CAG TCC CTG GTG GGC GGC CAG CAC GGC CAG ILe Tyr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ris Gly Arg AGC AAG GGG ACG GGC ACG GCC ACG CTG CTG GGC GGC CAG CAC CAC	
OTO GTO GGC CTG CTG GGC TAC GTG GTC TGG AGG TTC GCC CAG GCC TTC GTG GAC ACG GAG Val Val Gly Leu Leu Gly Tyr Val Val Trp Arg Phe Ala Gln Ala Phe Val Asp Thr Glu GAC AAG GGC TCC GAC GGC AAG GGC AAG GGC AAG GGC TCC ASP Lys Gly Ser Asp Ala Lys Gly Ile Ala Thr Arg Ala Met Tyr Phe Leu Ser Gly Cys ATC TAC GGG TCG CTG GCC GCC GCC GCC GCC GCC GGC GG	300 100
Val Val 31y Leu Leu Gly Tyr Val Val Trp Arg Phe Ala Gln Ala Phe Val Asp Thr Glu GAC AAG GGC TCC GAC GCG AAG GGC ACG GGC ACG GCC ATG TAC TTC CTC AGC GGC TGC ASP Lys Gly Ser Asp Ala Lys Gly 11e Ala Thr Arg Ala Met Tyr Phe Leu Ser Gly Cys ATC TAC GCG TCG CTG GCG TTC TTC GCC GCG CAG TCC CTG GTG GGC GCG GCG CAG CAG CTG TAC TAC GCG TCG CTG GCC TTC TTC GCC GCG CAG TCC CTG GTG GGC GCC GCG CAG TCC TYr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ris Gly Arg AGC AAG GGG AAG GCG TGG ACG CTG ATG GAG CAG CCC TTT GGC CGC GTG CTG Ser Lys Gly Thr Gln Gly Trp Thr Ala Thr Leu Met Glu Gln Pro Phe Gly Arg Val Leu GTG GGC CTG GTG GGC ATC GTG GGC TTC GCC CTG AAG CAG TTC CAC ACC GCG TGG Val Ala Leu Val Gly Leu Gly Ile Val Gly Phe Ala Leu Lys Gln Phe His Thr Ala TTP AAG GGG AAG TTC CGG GAG AAG CTC ACC CTC ACC GGA CTG GCT GCC CGG AAG CAG CAC CAC Lys Ala Lys Phe Arg Glu Lys Leu Thr Leu Thr Gly Leu Ala Ala Arg Lys Gln Ris Ris ATC GAG GCC ATG TCC CAG ATC GCC GGC GGC GGC GTG GTG TTC GCC GTC ATC GGC Ile Glu Arg Met Cys Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly GGC TTC CTC GTC CGC TCC GCC GTG GAC GCG GGC GTG GTG GTG GTG GCG GCG GTG GT	_
CAC ANG GGC TCC GAC GCG ANG GGA ATC GCC ACG CGC GCC ATG TAC TTC CTC AGC GGC TGC ASP Lys Gly Ser Asp Ala Lys Gly Ile Ala Thr Arg Ala Met Tyr Phe Leu Ser Gly Cys ATC TAC GCG TCG CTG GCG TCG GCG CTG GCG CTG GCG CTG GCG CTG GCG CTG GCG CTG GCG GC	360 120
ASP LYS Gly Ser ASP Als LYS Gly Ile Ala Thr Arg Ala Met Tyr Phe Leu Ser Gly Cys ATC TAC GCG TCG CTG GCC TTC TTC GCC GCG CAG TCC CTG GTG GGC GCC GCG CAC GGC CGG Ile Tyr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ris Gly Arg ACC AAG GGG ACG CAG GGC TGG ACG GCC ACG CTG ATG GAG CAG CCC TTT GCC CGC GTG CTC Ser Lys Gly Thr Gln Gly Trp Thr Ala Thr Leu Met Glu Gln Pro Phe Gly Arg Val Leu GTG GCG CTG GTG GGC ATC GTG GGC ATC GTG GGC TTC GCG CTG AAG CAG TTC CAC ACC GCG TGG Val Ala Leu Val Gly Leu Gly Ile Val Gly Phe Ala Leu Lys Gln Phe His Thr Ala TTP AAG GCG AAG TTC CGG GAG AAG CTC ACC CTC ACC GGA CTG GCT GCC CGG AAG CAG CAC CAC Lys Ala Lys Phe Arg Glu Lys Leu Thr Leu Thr Gly Leu Ala Ala Arg Lys Gln His His ATC GAG CCC ATG TCC CAG TTC GGC ATC GCC GCG GGC GTG GTG TTC GCC GTC ATC GGC Ile Glu Arg Met Cys Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly GGC TTC CTC GTC CGC TCC GCC GTG GAC GCG AAC CCC GGC GTG GTG GTG GCC CTG GGA GAG GCC CTG GCC GTC GTC GCG AGG CAG CCG TCC GGC GAC GTG CTC CTG GGG GTG GTG GCG GCG Ala Leu Ala Val Val Ala Arg Gln PTO Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GCC GTC GCC GCC TAC CCC TAC CTG TTC CTC CGC GCC CCC GAA CTC CTC GGC Ala Leu Ala Val Val Ala Arg Gln PTO Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GTG GCC TAC GCC GCC TAC CTG CTG CTG CTG CCC CCC GAA CTC CTC GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CCC TAC CCC GAA CTC TAG GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CCC TAC CCC GAA CTC TAG GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CCC TAC CCC GAA CTC TAG GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CCC TAC CCC GAA CTC TAG GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CCC TAC CCC GAA Ala Leu Ala Val Ala Arg Gln PTO Ser Gly Asp Val Leu Leu Leu Cly Val Val Ala Ala GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CCC TAC CCC GAA CTG CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CCC TAC CCC GAA CTG CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CCC TAC CCC GAA CTG CTG GTG GCC TAC	
ATC TAC GCG TCG CTG GCC TTC TTC GCC GCG CAG TCC CTG GTG GGC GCG CAG CAC GGC CGG Ile Tyr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ris Gly Arg AGC AAG GGG ACG CAG GGC TGG ACG GCC ACG CTG ATG GAG CAG CCC TTT GGC CGC GTG CTG Ser Lya Gly Thr Gln Gly Trp Thr Ala Thr Leu Met Glu Gln Pro Phe Gly Arg Val Leu GTG GGG CTG GTG GGG CTG GGC ATC GTG GGC TTC GCG CTG AAG CAG TTC CAC ACC GCG TGG Val Ala Leu Val Gly Leu Gly Ile Val Gly Phe Ala Leu Lya Gln Phe His Thr Ala TTp AAG GGG AAG TTC CGG GAG AAG CTC ACC CTC ACC GGA CTG GCT GCC CGG AAG CAG CAC CAC Lya Ala Lya Phe Arg Glu Lya Leu Thr Leu Thr Gly Leu Ala Ala Arg Lya Gln His His ATC GAG CGC ATG TGC CAG TTC GGC ATC GCC GGC GGC GGC GTG GTG TTC GCC GTC ATC GGC Ile Glu Arg Met Cya Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly GGC TTC CTC GTC CGC TCC GCC GTG GAC GCG AAC CCC GGC GAG GCC AAG GGC CTG GGA CAG GCC TTC CTC GTC CGC TCC GCC GTG GAC GCG AAC CCC GGC GAG GCC AAG GGC CTG GGA CAG GCC CTG GCC GTC GTC GCC AGG CAG CCG TCC GGC GAC GTG CTC CTC GGG GTG GTG GCG GCG Ala Leu Ala Val Ala Arg Gln Pro Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GCC TAC GCC TAC GCC TAC CTC TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG	420
The Tyr Ala Ser Leu Ala Phe Phe Ala Ala Gln Ser Leu Val Gly Ala Ala Ala Gly Arg ACC AAG GGG ACG CAG GGC TGG ACG GCC ACG CTG ATG GAG CAG CCC TTT GGC CGC GTG CTG Ser Lya Gly Thr Gln Gly Trp Thr Ala Thr Leu Met Glu Gln Pro Phe Gly Arg Val Leu OTG GCG CTG GTG GGG CTG GGC ATC GTG GGC TTC GCG CTG AAG CAG TTC CAC ACC GCG TGG Val Ala Leu Val Gly Leu Gly Ile Val Gly Phe Ala Leu Lya Gln Phe His Thr Ala Trp AAG GCG AAG TTC CGG GAG AAG CTC ACC CTC ACC GGA CTG GCT GCC CGG AAG CAG CAC CAC Lya Ala Lya Pha Arg Glu Lya Leu Thr Leu Thr Gly Leu Ala Ala Arg Lya Gln His His ATC GAG CGC ATG TGC CAG TTC GGC ATC GCC GCG GGC GTG GTG TTC GCC GTC ATC GGC Ile Glu Arg Met Cya Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly GGC TTC CTC GTC CGC TCC GCC GTG GAC GCG AAC CCC GGC GAG GCC AAG GGC CTG GGA GAG Gly Pha Leu Val Arg Ser Ala Val Asp Ala Asn PTO Gly Glu Ala Lya Gly Leu Gly Glu GCC CTG GCC GTC GTC GCG AGG CAG CCG TCC GGC GAC GTG CTC CTG GGG GTG GTG GCG GCG Ala Leu Ala Val Val Ala Arg Gln PTO Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC GAA CTC TAC GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GCC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG	140
ACC AAG GGG ACG CAG GGC TGG ACG GCC ACG CTG ATG GAG CAG CCC TTT GGC CGC GTG CTG Ser Lym Gly Thr Gln Gly Trp Thr Alm Thr Leu Met Glu Gln Pro Phe Gly Arg Val Leu GTG GGC CTG GTG GGG CTG GGC ATC GTG GGC TTC GCC CTG AAG CAG TTC CAC ACC GCG TGG Val Alm Leu Val Gly Leu Gly Ile Val Gly Phe Alm Leu Lym Gln Phe Him Thr Alm TTp AAG GCG AAG TTC CGG GAG AAG CTC ACC CTC ACC GGA CTG GCT GCC CGG AAG CAG CAC CAC Lym Alm Lym Phe Arg Glu Lym Leu Thr Leu Thr Gly Leu Alm Alm Arg Lym Gln Him Him ATC GAG CGC ATG TGC CAG TTC GGC ATC GCC GCG GGC GTG GTG GTG GCC GTC ATC GGC Lle Glu Arg Met Cym Gln Phe Gly Ile Alm Alm Arg Gly Val Val Phe Alm Val Ile Gly GGC TTC CTC GTC CGC TCC GCC GTG GAC GCG AAC CCC GGC GAG GCC AAG GGC CTG GGA CAG Gly Phm Leu Val Arg Ser Alm Val Amp Alm Amn Pto Gly Glu Alm Lym Gly Leu Gly Glu GGC CTG GCC GTC GTC GCC AGG CAG CCG TCC GGC GAC GTG CTC CTG GGG GTG GTG GCG GCG Alm Leu Alm Val Alm Arg Gln Pto Ser Gly Amp Val Leu Leu Gly Val Val Alm Alm GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC GAA CTC TAG GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG	480
SET LYB GLY THE GLE GLY TEP THE ALB THE LEW MET GLU GLE PEO PRO GLY ARG VAL LEW OTO GCG CTG GGG CTG GGC ATC GTG GGC TTC GCG CTG AAG CAG TTC CAC ACC GCG TGG VAL ALB LEW VAL GLY LOW GLY THE VAL GLY PRO ALB LEW LYB GLE PRO CAG CAC CAC CAC LYB ALB LYB PRO ARG GLG AAG CTC ACC CTC ACC GGA CTG GCT GCC CGG AAG CAG CAC CAC LYB ALB LYB PRO ARG GLU LYB LEW THE LEW THE GLY LOW ALB ALB ARG LYB GLE HIS BLIS ATC GAG CAC CAC GCC ATC GCC GCC GCC GTG GTG TTC GCC GTC ATC GCC LIG GLU ARG MET CYB GLE PRO GLY LIE ALB ALB ARG GLY VAL VAL PRO ALB VAL LIE GLY GGC TTC CTC GTC CCC GCC GTG GAC GCC GAG GCC AAG GCC CTG GGA GAG GLY PRO LEW VAL ARG SET ALB VAL ABG ALB ARG PTO GLY GLU ALB LYB GLY LEW GLY GLU GLY GLU ALB VAL ARG GLC CTG GCC GTC GCC GCC GAC GTG CTC CTC GCC GCG GCG AAG GCC CTG GGG GTG GTG GCC GCG AAB ACC CTC GCC CTC GCC GCC GCC GTG GTG GTG G	160
OTG GCC CTC GTC GGC GGC ATC GGC ATC GTC GGC TTC GCC CTG AAC CAG TTC CAC ACC GCC TOG Val Ala Leu Val Gly Leu Gly Ile Val Gly Phe Ala Leu Lys Gln Phe His Thr Ala TTP AMG GCC AAG TTC CGC GAG AAG CAC CTC ACC CTC ACC GGA CTG GCT GCC CGG AAG CAG CAC CAC Lys Ala Lys Phe Arg Glu Lys Leu Thr Leu Thr Gly Leu Ala Ala Arg Lys Gln His His ATC GAG CGC ATG TCC CAG TTC GGC GTC GCC GGC GTC GTC GTC GTC ATC GGC Ile Glu Arg Met Cys Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly GGC TTC CTC GTC GGC TCC GCC GTG GAC GCC AAC GCC AAG GGC AAG GGC CTG GGA GAG GCC AAG GGC CTG GGA GAG GCC AAG GCC CTG GGA GAG GCC CTG GCC GCC GCC GTC GCC GCC GAC GTC TCC GCC GCC GCC GCC GAC GCC AAG GCC AAG GCC AAG GCC AAG GCC CTG GCC GCC GCC GCC GCC GCC GCC GC	540
Val Ala Leu Val Gly Leu Gly Ile Val Gly Phe Ala Leu Lys Gln Phe His Thr Ala TTP ANG GCG ANG TTC CGG GAG ANG CTC ACC CTC ACC GGA CTG GCT GCC CGG ANG CAC CAC Lys Ala Lys Phe Arg Glu Lys Leu Thr Leu Thr Gly Leu Ala Ala Arg Lys Gln His Bis ATC GAG CGC ATG TGC CAG TTC GGC ATC GCC GCG GGC GTG GTG TTC GCC GTC ATC GGC Ile Glu Arg Het Cys Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly GGC TTC CTC GTC CGC TCC GCC GTG GAC GCG AAC CCC GGC GAG GCC AAG GGC CTG GGA GAA Gly Phe Leu Val Arg Ser Als Val Asp Ala Asn PTO Gly Glu Ala Lys Gly Leu Gly Glu GCC CTG GCC GTC GTC GCG AGG CAG CCG TCC GGC GAC GTG CTC CTG GGG GTG GTG GCG Ala Leu Als Val Val Ala Arg Gln PTO Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC GAC GTG CTC GAA CTC TAG	180
ANG GOG ANG TTC CGG GAG AAG CTC ACC CTC ACC GGA CTG GCT GCC CGG ANG CAG CAC CAC Lys Als Lys Pha Arg Glu Lys Leu Thr Leu Thr Gly Leu Als Als Arg Lys Gln His His Atc GAG CAC CAC CAC GGA CGC GGC GTG GTG TTC GCC GTC ATC GGC Ile Glu Arg Met Cys Gln Pha Gly Ile Als Als Arg Gly Val Val Pha Als Val Ile Gly GGC TTC CTC GTC GCC GTG GTG GAC GCG AAC CCC GGC GAG GCC AAG GGC CTG GGA GAG GLY Pha Leu Val Arg Ser Als Val Asp Als Asn PTO Gly Glu Als Lys Gly Leu Gly Glu GCC CTG GCC GTC GCC GCC GAC GCC GAC GTG CTC GCC GCC GCC GAG GCC AAG GCC AAG GCC AAG GCC CTG GCC GCC GCC GCC GCC GCC GCC GC	600
Lys Ala Lys Phe Arg Glu Lys Leu Thr Leu Thr Gly Leu Ala Ala Arg Lys Gln His His Arc GAS CCC Arg TGC CAG TTC GGC ATC GCC GCC GGC GGC GGC GTG GTG TTC GCC GTC ATC GGC Lle Glu Arg Met Cys Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly GGC TTC CTC GTC CGC GCC GTC GGC GAC GCC GAG GGC AAG GGC CTG GGA GAG GLy Phe Leu Val Arg Ser Als Val Asp Ala Asn Pro Gly Glu Ala Lys Gly Leu Gly Glu GCC CTG GCC GTC GTC GCG GAG CAG CCG GAC GTG CTC CTG GGG GTG GTG GCG GCG Ala Leu Ala Val Ala Arg Gln Pro Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GTG GCC TAC GCC GCC TAC GCC GCC TAC GCC GAC GTG CTC CAG GCG GAA CTC TAG	200
ATC GAG CGC ATG TGC CAG TTC GGC ATC GCC CGC GGC GGC GTG GTG TTC GGC GTC ATC GGC Ile Glu Arg Het Cys Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly GGC TTC CTC GTC GCC GTG GAC GGC GAG GCC AAG GGC CTG GGA GAA GAA GAY Phe Leu Val Arg Ser Als Val Asp Ala Asn PTO Gly Glu Ala Lys Gly Leu Gly Glu GCC CTG GCC GTC GTC GCG GAG GCC GAC GTG CTC CTG GGG GTG GTG GCG GCG Ala Leu Als Val Ala Arg Gln PTO Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GTG GCC GTC GCC GCC TAC GCC GCC TAC GCC GAA CTC TAG	660
Ile Glu Arg Met Cys Gln Phe Gly Ile Ala Ala Arg Gly Val Val Phe Ala Val Ile Gly GGC TTC CTC GTC CGC TCC GCC GTG GAC GCG AAC CCC GGC GAG GCC AAG GGC CTG GGA GAG Gly Phe Leu Val Arg Ser Ala Val Asp Ala Asn Pro Gly Glu Ala Lys Gly Leu Gly Glu GCC CTG GCC GTC GTC GCG AGG CAG CCG TCC GGC GAC GTG CTC CTG GGG GTG GTG GCG GCG Ala Leu Ala Val Val Ala Arg Gln Pro Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG	220
GOT TTC CTC GTC CGC TCC GCC GTG GAC GCG AAC CCC GGC GAG GCC AAG GGC CTG GGA GAG Gly Phe Leu Val Arg Ser Ale Val Asp Ale Asn Pro Gly Glu Ale Lys Gly Leu Gly Glu GCC CTG GCC GTC GTC GCG AGG CAG CCG TCC GGC GAC GTG CTC CTG GGG GTG GTG GCG GCG Ale Leu Ale Val Val Ale Arg Gln Pro Ser Gly Asp Val Leu Leu Gly Val Val Ale Ale GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC GGC GAA CTC TAG	720
Gly Phe Leu Val Arg Ser Ale Val Asp Ale Ann Pro Gly Glu Ale Lys Gly Leu Gly Glu GCC CTG GCC GTC GTC GCG AGG CAG CCG TCC GGC GAC GTG CTC CTG GGG GTG GTG GCG GCG Ale Leu Ale Val Val Ale Arg Gln Pro Ser Gly Asp Val Leu Leu Gly Val Val Ale Ale GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC GGC GAA CTC TAG	240
GCC CTG GCC GTC GCC GCC AGG CAG CCG TCC GGC GAC GTG CTC CTG GGG GTG GTG GCG GCG Ala Leu Ala Val Val Ala Arg Glm Pro Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC GGC GAA CTC TAG	780
Ala Leu Ala Val Val Ala Arg Glm Pro Ser Gly Asp Val Leu Leu Gly Val Val Ala Ala GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG	260
GGC CTG GTG GCC TAC GCC GCC TAC CTG TTC CTC CAG GCG CGC TAC CGC GAA CTC TAG	840
	280
Gly Leu Val Ala Tyr Ala Ala Tyr Leu Phe Leu Gln Ala Arg Tyr Arg Glu Leu End	897
	299

Pigure 16

Whale Not Sample 11.801 Esterase es2

AT	3 AC	C AA r Ly	T A	rc G	CA la	ATA Ile	Leu	TI	GCG Ala	Lev	ATI	A AO	G GC	A TA a Ty	C CT	G CC	ون دي وي	A C	CT G	TG al	ATG	. 6
AA	cre	3 OT	A 77	T T	TA	GGG	CCC	CGC	GAR	ACC	CT.	GG	3 GC	A 03	G AC	ст	7 (37	c o	TA	AA	GCC	12
												_			g Thi 3 GT(-				. 9: 18:
Gln	Ala	Va.	1 91	y A	79	Leu	Ala	Asn	Ala	Thr	Arg	Pro	Va.	l Oly	y Val	Ile	Pr	o Th	r V	11	Glu	66
GJ <i>n</i> G Y Y	AGC 9er	Are	Ly	G AT B Me	t	ACG Thr	CAT ABP	aaa Lyb	YIB	GII Val	AGC Ser	Leu	Phe	CA:	GIN CAG	Pro	A)	c cc	o (1)	LA Lu	TTA Leu	. 240 80
TTC Phe	CGT Arg	Lys	Ly	A GI B As	p :	ATT Ile	CAG Gln	ATT Ile	GAC Asp	GGO Gly	GCT Ala	GAA Glu	OGG Oly	CCI	Ile	GAT ABP	GC	C CG	T AI g Il	.e	TAC Tyr	300
AGC Ser	gyy gcc	CCI	A):	A AA A Ly	A (CAT Bis	CGC Arg	CCN Pro	CON Arg	CCA Pro	ATN 11e	CTA Leu	GTG Val	TAT	TIT	CAC	600	00	y G1	y :	IGG ITP	360 120
GTT Val	CAG Gln	GGC	RAS TBA	CT	9 E	rab Syc	AGC Ser	CAT His	GAC Asp	GGG Gly	GII Val	TGC Cys	GGC	AAG Lys	CIG	GCA Ala	Lye	TE	GC GC	2 ; a ;	AAC Aan	420 140
TGC Cys	ATT 1le	GII Val	ATC 11e	TC Se	0 G	TC (GAT Asp	TAT Tyr	CGT Arg	CTA Leu	GCG Ala	CCC Pro	GAA Glu	CAC His	AAA Lys	TTT Phe	CCI	Cyt	C GC	O C	OTO	480 160
CIT Leu	GAT Asp	GCG GCG	ATI Ile	GC Al:	3 G	icc :	TAT Tyr	AAA Lys	11D 1GO	GTG Val	CGC Arg	GCC Ala	AAC Aøn	GCA Ala	ACA Thr	AAC Asn	CIT	GGC	GG GL;	2 G	AT sp	540 180
CCT Pro	GAA G1 <i>u</i>	CGT Arg	ATC	G1;	2 G 7 V	al (GGC 31y	GGC GGC	GAT Asp	AGC Ser	GCA Ala	GCG GLy	GJA GGC	AAT ABN	CTT Leu	GCC Ala	GCC Ala	GIT Val	Val	1	GC Ya	600 200
ULD CVY	CAA Gln	ACC Thr	GCC	Met	A E	AC (3GC (GRG Glu	CGC : Arg	ACA Thr	CCA Pro	CAT Asp	CTG CTG	CAA Gln	GTC Val	CTG Leu	ATC 11e	TAT Tyx	CCC	G	00 la	660 220
CTG	GAT Asp	GCA Ala	CGC Arg	ATC	A	TC 1 le 5	er :	ACC The	TCG / Ser !	ATG (GAG 31u	GAA Glu	TTG Lev	CGT Arg	GAT Asp	GCC Ala	TAC Tyr	ATC 11e	TTG	P	50	720 240
															CCI Pro							780- 260
eu .	AGG Arg	DTA Jet	TCG Ser	Pro	A:	IT C	TC /	red (IAT / LSp 1	ter (TC (/al)	3C0 (TAE QeA	CAA Gln	CCC Pro	CAA Gln	GCC Ala	TGC Cys	ATT 11e	G7 Va	1C	840 280
															CGA A							900 300
	ABP			TA									•									914 304

Pigure 17

Whale Mat Sample AD3059 Seterase es4

GT(AGG	AT	T CG	r cro	CO.	CTO	TT	AAC	1G0	TT	TI	G AA	r AC	C TT	CAD	AA	A CC	A AA	A CTG B Leu	. 2
																_		•		
Ala	Ala	Al	Ly	a Acc	Pro	Asp	Asp	Leu	Arg	Lys	Sei	Phe	Gli	a Leu	Lys	A)	A ALT) Ph	TTG Leu	12
TT1 Phe	Pro	GCC Ali	CCI Pro	A CG1	Lys	ACA Thr	AGG	Phe	AGI	CAT His	GA:	val	Lei	G CAG	TCA Ser	GGC	TATO	GGC	TCG	18: 6:
GTA Val	AAT	GCC	CAC	TGG	GCG	AAA	TCC	AAA	TCI	GCA Ala	TCI	GAT	GAC	AGG	GTA Val	ATC	CTO	TAT	TIT Phe	246
CAT	GGG	GGZ	GGG	TAT	GII	TIT	GGG	TCA	CCA	AAA	ACS	CAC	: CG1	r GCA	DTA	TTG	GCG	CGC	TIG	300
TCG	GCA	ATG	ACA	GGT	CTI	ICI	GCO	TGC	CTT	CCA	GAT	TAT	AGG	770	GCA	CCA	GAG	CAC	CCA	360
Ser	Ala	Met	Thr	Gly	Leu	Ser	Ala	Сув	Leu	Pro	Yab	Tyr	Arg	Leu	Ala	Pro	Glu	His	Pro	120
																			ATC	420 140
				ATT Ile															TTG Leu	480 160
				AAG Lys																540 180
				TCA Ser																600 200
				CGG Arg																660 220
IAT	CCA	CCT	GCA.	TCG Ser	ccs	CTG	CAG	cca	CAT	TTT	TCT	GGC	ATG	ccs	CCT	GTA	TTT.	CTG	ACA	720 240
iCA.	AGT	GAC	agt	GAA Glu	ATC	CIG	TTG	GAT	CAT	TGC	CZG	CGG	ATG	CCC	gat	CAC	TTG	œı	GCG	780 260
		-		GTG .				•	•	•		•	-		•			_		840
				Val																280
				Pro																900 300
				TCA . Ser .				TA												308 356

Picrore 18

Microscilla furvescens Esterase 53sc2

																			ATG Met	6
III	TTA	GGG	CTA	TCA	GCA	(COA	GAG	GCT	CAA	GAC	TGG	CCI	GAC	: CTA	CAG	. AAA	TAI	. C31	AGT	12
Pho	Leu	Gly	·· Leu	Ser	Ala	Ala	Glu	Ale	· Gln	Asp	IIP	Pro	Aep	Leu	Gln	Lys	zAz	. Arg	Ser	4
																			AAC	180
Ala	Asn	Lya	Glu	Ala	Lyø	Leu	Leu	Pro	Lys	G) <i>n</i>	GBA	Arg	Lys	Val	Val	Phe	Met	gly	Asa	6
			GAA																	240
Ser	Ije	Thr	Glu	Ala	đz	Ile	Ser	Gln	Arg	Pro	Glu	Phe	Pho	Ser	Glu	Asn	Gly	Phe	Ile	80
			ATC																	300
Gly	Arg	Gly	Ile	Ser	Gly	Gln	Thr	Thr	Pro	G]n	Met	Leu	Leu	yrg	Phe	Arg	Gln	Asp	Val	100
			CAG																	360
Ile	Увр	Leu	Gla	Pro	Lys	Ala	Val	Val	Ile	Leu	Ala	Gly	Thr	Asn	Авр	Val	Ala	Gln	Asn	120
			DTA																	420
Thr	Gly	Pro	Met	Thr	Ile	Glu	Glu	Ser	Leu	Ala	Asn	Ile	Lys	9er	Met	Val	Glu	Leu	Ma	140
			GGG																	480
3ln	Ala	Asn	Gly	Ile	Thr	Pro	Val	Leu	Суз	Thr	·Val	Leu	Pro	Ala	Asp	Arg	Phe	Ser	Trp	160
CGA	CCI	GN0	CTT	ACA	ccc	GCA	CAA	ACT	ATC	ATT	GCC	CTC	aat	CAG	CLC	ATT	AAG	CAA	TAT	5,40
/xg	Pro	Glu	Leu	Thr	Pro	Ala	Glu	Thr	Ile	Ile	Ala	Leu	Asn	Gln	Leu	Ile	Lys	Gln	Tyr	180
SCC	GAG	GCA	C)G	GCC	CTG	ecc	CTG	C10	CAT	TAT	CAT	GCT	GCA	CIC	YCC	aat	AAA	GGT	GGA	600
119	GIA	ATS	Gln	GIA	ren	ATB	ren	vai	Asp	TYT	H78	ATS	Ala	Leu	Thr	ABD	Lys	gly	GIA	200
			GIG																	660
ily	Leu	Pro	Val	ьyв	IYE	OTA	GT.	ABP	GIA	AST	HTS	PTO	ABD	ABI	ATG.	GIY	TYT	arv	AQT	220
			ATT														TA			713
re C	GIU	WRU	lle	491	Let'U	PIO	AHT	TTE	oe F	Jer	GIU	Ten.	AL3	PAR	ia:U	ьyв				237

Pigure 19

Thermotoga maritima MSB8 Esterase 6scl

																				•
AT Ne	G GC	C TI a. Ph	C TI	C GA e As	T TI p Le	A CC	A CTO	Gl Gl	GA GA	A CT	G AA 1 Ly	G AA S Ly	A TA B Ty	T CG	r cc	A GA	G CG	G TA g Ty	C GN T Glu	L 60
GA G1:	G AA	A GA	C TI p Ph	C GA e As	T GA	G TT	C TGC	GAI Glu	GA	G AC	A CT	c GC	A GA	G AG	GAL Glu	A AAI	3 TI 9 Ph	C CC	C TTA	1 22
GA	ב ככו	c GT	C TT	C GA	G AG	G ATC	GAG	TCI	CAC	: CT	. AA	A AC	A GT	c gaz	GCC	TAC	CA	T GT	C ACC	180
																			l Thr	•
Pho	s Sei	c GT	y Ty	r Ari	3 GG1	y Glr	AGG Arg	Ile	Lys	GG GI	Tr	Len	Le	r GT1	Pro	LYS	Le	G GA	n Gyn	240 80
CAI	AAJ	CT	CC	TG	GT1	GTG	CAG	TAC	ATA	GCA	TAC	AAC	GGT	CGA	AGA	GG	110	: cc	CAC	300
																			BiH c	100
GA(As ₁	TIT	CTO Len	3 TT	TG(S CCI	TCT Ser	ATG Met	GCT	TAC	ATA Ile	Cys	Phe	GTC Val	ATG Met	GAT Asp	ACI Thr	CG!	(G1)	CAG	360 120
GG#	AGC Ser	GG(Tri	CTC	AAA Lys	GGA	GAC	ACA	CCG	GAT	TAC	CCT	GAG Glu	GGT	CCC	GII Val	GAC	CC1	CAG Gln	420 140
_			_			_	_			-	_						_		GTC.	480
Tyr	Pro	Gly	Phe	Met	Thr	Arg	Gly	Ile	Leu	Asp	Pro	Arg	Thr	Tyr	Tyr	Tyr	Arg	Arg	Val	160
TTC Phe	ACG Thr	CAC	GCT Ala	GTC Val	AGA Arg	SCC Ala	GTT Val	gaa Glu	GCT Ala	GCT Ala	GCT Ala	TCT Ser	TTT Phe	Pro	CAG Gln	GTA Val	gat Abp	Gyu CYY	GLU	540 180
														CTT						600
									_	-	•			Leu						200
Ser	Lys	Lys	GCA Ala	Lys	Ala	Leu	Leu	CAB	QAT Asp	Val	Pro	Phe	Leu	CAR JAJ.	His	Phe	AGA	YCY	GCA Ala	660 220
TTA Val	CAG Gln	CTT Leu	GTG Val	gat Asp	ACG Thr	CAT His	CCA PTO	TAC Tyr	GCG Ala	GAG Glu	ATC 11e	ACG Thr	AAC Asn	TTT	CTA Leu	aag Lys	ACC Thr	CAC His	AGA Arg	720 240
3AC	AAG	GAA	GAA	ATC	GTG	TIC	AGG	ACT	CTT	TCC	TAT	TTC	GAT	GGA	GTG	AAC	TTC	GCA	GCC	780
_	=						_				•		-	Gly						260 ;
ra Ta	GCG Ala	aag Lys	ATC Ile	CCI Pro	YJ9 GCC	CTG Leu	TTT Phe	TCT Ser	GTG Val	GCT	CTC Leu	ATG Met	GAC ABP	AAC Aan	ATT Ile	Cys TGT	Pro Pro	CCT PTO	TCA Ser	840 280
LCG thur	GTT Val	TTC Phe	GCT	GCC Ala	TAC	AAT	TAC	TAC	GCT Ala	GGA G1 v	CCG Pro	AAG Lova	GAA	ATC . Ile .	AGA	ATC	TAT	cca	TAC	900 300
					-		-	-		-		-		GTG :	_		-		•	960
														Val						320
	TTT Phe				TAA															978 326

Figure 20

Polyangium brachysporum Esterase 78mcl

TTG	AAG	TAC	: 110	' AAA	CCC	: ccc	CII	. CCC	GGC	ATC	ACC	110	CIC	: GGC	CTG	CTG	GCC	TGC	ACC	. 6
teu	Lva	TVX	Phe	Lve	Ala	Arg	Leu	Ala	Glv	Ile	Thr	Leu	Leu	Glv	Leu	Leu	Ala	CVS	·Thr	. 20
	-,-	-,-																		
***	acc	T(1)	GCG	CAG	200	GAG	ccc	ATC	GTR	TT	12772	Cac	aac	TAT	TOO	GCC.	ייטע	CCA	TCC	120
																		Ala		40
Ser	A10	· >er	VIO	GIII	141	610	FLU	116	Val	. PHG	Vaj	ars	GLY	TAE	361	GLY	941	MIG	JEL	. 4
	~~~	920	300	370	~~	oac	~~	-	cca	***	220	^~	***	~~	***	000	~~	CTC	TAC	180
																				60
ABR	11D	ABD	ınr	wec	ren	GLY	w.a	rne	Arg	ser	ABN	GIA	TYE	Ala	ser	GIY	ser	Leu	TYP	. •
			<b>*</b>	***	****	-	_	300	200		000		300	~~~	100	CNO	~~	CGC	200	240
Thr	Phe	ASD	TYX	ABN	ser	Leu	AST	ser	ser	ABD	Arg	Thr	Ser	ALA	Ser	GTA	Leu	Arg	ser	80
				~~~	~~	~~	~~~	636	~~~		~~~	~~~			~~~	~~~	~~~	CAC	***	300
Phe	Val	ABA	TAT	ABT	Arg	ser	Arg	HIS	GIY	ASD	ATB	Arg	ITE	A19	ren	VAI	YTP	His	ser	100
				_		~~~	-		~~~		~~~		~~~	~~~	~~~	.~	~~~	ACC	~~	360
Asn	GIA	grà	Leu	ABI	ser	Arg	ub	IJI	Arg	ATS	GIU	Leu	GIA	GIA	GIG	TAE	ALB	Thr	Arg	120
																				420
																		TAC		
Arg	Pne	Val	TILE	Leu	GTA	Inr	NIO	H18	vra	ary	INE	THE	11b	ALA	TYE	ATB	CAB	IYI	ser	140
				~~~		~~~	~~~	~~~	-~			~~~	100	.~~	~~~	~~~		CGT	~~~	480
Pro	Ala	Cys	Phe	GIA	Mec	Arg	PIO	GTA	ser	Ser	ren	ren	IDI	TDE	ren	ату	Ser	Arg	ATB	160
				~~~		~~		. ~~	C . C	~~~			~	~~	^~	***	100		ca c	540
TOC	GAC	CGC	TCG	CIG	100	ICA	MAC	ACC	GML	المال	Wif	WIC	CIG	CUG	oc.	100	WOL.	GCG		
CAB	Asp	Arg	Ser	Leu	TTP	Ser	Aen	Thr	qsa	GIA	Ile	Ile	Leu	Pro	YTF	ser	ser	Ala	GIL	180
																			~~~	600
TGT	GGT	GIC	AGC	ACG	CGC	ACT	GCC	GAL	GIC	AGC	CAT	CIC	GAC	CIU	CIG	ACC	GAC	TCI	CGC	
Cys	Gly	Val	Ser	Thr	Arg	Thr	ALB	ASP	VAI	Ser	His	Leu	ysb	Leu	Leu	The	Asp	Ser	Arg	200
					~~~				~~~							<i>-</i>		~~~	~~~	
GIG	TAC	ACG	CAB	T.U	COC	ACG		LIG	CAA	TUA	Circ	IVA	CGG	TUC	ACC	GAA.	CUT	GCA	CCT	660
Val	Tyr	TUL	alb	Leu	AFG	Thr	GIU	ren	GTD	Bnd	gly	KDd	Arg	Cys	TOF	ara	Arg	Ala	PXO	220
_																				661
G																				22T